COST
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

COST Action CA20112

COOPERATION WITH THE REVALUATION CONFERENCE

WWW.PROFEEDBACK.EU

PROFEED book8

EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

This publication is based upon work from COST Action CA20112 PROFEEDBACK, supported by COST (European Cooperation in Science and Technology). COST (European Cooperation in Science and Technology) is a funding agency for research and innovation networks. Our Actions help connect research initiatives across Europe and enable scientists to grow their ideas by sharing them with their peers. This boosts their research, career and innovation. Visit www.cost.eu

WWW.COST.EU

Funded by the European Union

PROFEEDBACK

CA20112 - PlatfoRm OF policy Evaluation community for improvED EU policies and Better ACKnowledgement

This publication is based upon work from COST Action CA20112 PROFEEDBACK, supported by COST (European Cooperation in Science and Technology). COST (European Cooperation in Science and Technology) is a funding agency for research and innovation networks. Our Actions help connect research initiatives across Europe and enable scientists to grow their ideas by sharing them with their peers. This boosts their research, career and innovation. Visit www.cost.eu

ISBN: 978-615-6756-10-7

Visit: www.profeedback.eu

PROFEEDBACK

CA20112 - PlatfoRm OF policy Evaluation community for improvED EU policies and Better ACKnowledgement

PUBLISHER

HETFA Research Institute
Grant Holder Institution
www.hetfa.eu

ACTION CONTACTS

Gábor BALÁS
Action Chair, Scientific Representative
profeedback@hetfa.hu

Tarmo KALVET
Action Vice Chair
tarmo.kalvet@ttu.ee

Renata Anna JAKSA
Science Communication Coordinator
rajaksa@hetfa.hu

Dijana ŠTRBAC
Innovation and Exploitation Manager
dijana.strbac@pupin.rs

Hanife AKAR
Grant Awarding Coordinator
hanif@metu.edu.tr

Anna SZENTES
Grant Holder Manager
profeedback@hetfa.hu

TABLE OF CONTENTS

ABOUT PROFEEDBACK

The COST Action PROFEEDBACK - PlatfoRm OF policy Evaluation community for improvED EU policies and Better ACKnowledgement (CA20112, MoU 052/21) 2021-2025 aims to foster the networking of the policy evaluation community at EU-level, raise awareness on the importance of evaluation policy research and improve its impact on policy-making. The Platform, following a bottom-up and open approach, **gathers** researchers and professionals from various scientific fields and sectors to present and evaluate theories, topics, tools and methods of policy evaluation. Results of the Europe-wide assessment of good practices **provides** direct and high-quality inputs for national and EU bodies responsible for policy evaluation. Policy evaluation is a key tool in understanding, developing and modernising EU policies, thus there is a growing demand for EU-wide and high quality evaluation services.

The main challenge is the shortage of sufficient bottom-up platforms for European researchers and professionals working in policy evaluation. They have limited possibilities to discuss common problems, assess country specific practices and share their knowledge in a mutually beneficial and effective way. The PROFEEDBACK Action aims **to contribute to these discussions during its four Grant Periods**, 18/10/2021 - 17/10/2025 and beyond.

The PROFEEDBACK Action has three key objectives to achieve during the Action period 18/10/2021 - 17/10/2025 and beyond.

Foster networking and knowledge-exchange of the policy evaluation community at European level

Raise awareness on the importance of policy evaluation research and improve its impact on policy-making

Reinforce state-of-the-art research in the policy evaluation field and contribute to evaluation standards

ABOUT PROFEEDbook8

In the framework of the PROFEEDBACK Action, one conference every half-year is implemented. After each conference thematic deliverables are issued, the so called PROFEEDbooks (D4.1-D4.8). They summarise the main presentations and results of the conferences.

PROFEEDbooks support multidisciplinarity, the systematisation of different methodologies and the exact transfer of know-how for the policy evaluation community. These also serve the interests of European and national policy-makers in developing the frameworks of the next programming periods and of a common European evaluation culture.

In addition, the PROFEEDbooks enrich the literature of public policy, aiming to develop theory, knowledge, method and tool base of European evaluation policy as well as a common understanding of the current problems and challenges. PROFEEDbooks support the Action in reaching audience beyond the Action members.

The **8th PROFEEDBACK** Conference is hosted in collaboration with the REvaluation Conference. This event centers on a**dvancing innovative and inclusive approaches to evaluating research, innovation, and policy**, with a strong emphasis on **interdisciplinary and cross-sectoral collaboration**.

Key sessions explore topics such as responsible research assessment, transformation policies, and participatory evaluation methods. Panels and workshops address emerging trends, including the impact of AI on funding and evaluation, strategies for measuring transformative outcomes, and insights into sustainable development practices. In addition, dedicated discussions focus on advancing evaluation practices within widening countries, supporting efforts to bridge evaluation gaps across diverse European contexts.
The conference provides a platform for in-depth dialogue and knowledge exchange, equipping evaluators and policymakers with strategies to navigate and adapt to evolving challenges in research and innovation assessments.

The event is organised in **Vienna, Austria**, from **December 4-6, 2024**.

The organisers are the **Austrian Institute of Technology** in collaboration with the **University of Vienna**.

Evaluation of Transformation Policies and Dynamics in Socio-technical Systems

Verena Régen, Brigitte Ecke
WPZ Research, Austria

EVALUATING TRANSFORMATIVE INNOVATIONPOLICY INSTRUMENTS

ABSTRACT

This contribution focuses on the evaluation of the program INNOVATORINNEN and seeks to provide practical insights into the evaluation of transformative innovation policies (TIP). The authors of the contribution were evaluators of the program INNOVATORINNEN in the years 2022 and 2023 and are currently conducting a small study at the onset of a new program strand (INNOVATORINNEN Lab), which has resulted from the evaluation. Preliminary evaluation results were presented at the Eu-SPRI conference 2023 (Régent, 2023). INNOVATORINNEN is a research promotion program specifically for female researchers and innovators that was initiated by the Austrian Ministry of Labour and Economic Affairs (BMAW) in 2022 and administered by the Austrian Research Promotion Agency (FFG). In early 2024, the program has started into its third round.

In contrast to more traditional programs for the promotion of women in science and research, INNOVATORINNEN explicitly takes an interdisciplinary and intersectional approach. It acknowledges that scientific careers are increasingly non-linear and often do not follow the “typical” academic path. The creation of new knowledge, developments, and products increasingly happens at the intersections of the research and innovation system: Many important impulses for tackling the Grand Challenges and working towards the SDGs come from areas that are not traditionally rooted within the academic sector. Next to female researchers that work in classical (university) fields of research and science, the program INNOVATORINNEN therefore addresses highly qualified women also from other sectors (e.g., education, creative industry, or the social sector). For the FFG, the program represents an entirely new approach: while it usually addresses institutions, INNOVATORINNEN is focused on the personal promotion of individual researchers with tools such as mentoring, training, and coaching (non-monetary support) aiming at yielding individuals and, as a long-term consequence, systemic impacts (Régent et al., 2023).

The authors of this contribution argue that INNOVATORINNEN is an example of a TIP instrument, TIP being an emerging generation of innovation policies, reorienting public science funders’ and innovation policy professionals’ efforts for initiating or contributing to societal change (Ghoshet al., 2021). Schot et al. (2019) define emerging TIPas as characterised by instruments that aim at fostering new connections between systems, providing spaces for experimentation and co-creating solutions for broader socio-technical system change. These characteristics can be found in the INNOVATORINNEN program: The continuing under-representation of women in leading roles in science and innovation is

an ongoing societal problem that disadvantages a major proportion of the population (e.g., Wroblewski, 2022; Greussing et al., 2016; OECD, 2016; Klapfer & Moser, 2022; Wisenöcker et al., 2021) and holds far-reaching consequences for society t large. For example, innovation risks to remain one-sided and to ignore non-male life realities (Marçal, 2021); moreover, recent studies suggest that if female scientists had more decisive power, they would more strongly pursue research projects to solve social and ecological problems and work towards changing work conditions and collaborative practices (Régent & Ecker, 2024). The INNOVATORINNEN program aims at tackling these issues, fostering connections between different sectors relevant to the innovation system, and providing spaces for experimentation and co-production of solutions for broader societal problems.

Within the FFG, INNOVATORINNEN is situated in the strategy department (much rather than in the classical funding administration), which experiments with new formats and target group-specific offerings. INNOVATORINNEN considers itself a learning program and uses its accompanying evaluation for being constantly informed about its participants feedback with a view to aspects related to the content and organisation of the program, but also regarding its immediate impacts on their personal and professional development (on an aggregated level).

As regards evaluation, as highlighted by Wise et al. (2022), the emergence of transformative innovation policy also creates new requirements for policy evaluation: Next to the traditional purposes of evaluation—assessing efficiency, effectiveness, and the relevance of policy programs (Peersman, 2015)—Boni et al. (2019) call for a new evaluation strategy that comprises monitoring progress and informing the direction of the pursued systemic change process. In line with Molas-Gallart et al. (2021), the authors stress the integration of evaluation as a strategic dimension of the given program with the aim to enhance reflexivity and learning. Based on Patton (2006), Chataway et al. (2017), and Molas-Gallart et al. (2020), Wise et al. (2022) have formulated the key characteristics of evaluations of transformative innovation policies (p. 274):

- Help to inform and refine the transformation process (directionality, societal goals, and system impact)
- Use a mix of methods and techniques to assess and contextualise transformative outcomes/signs of change
- The evaluation process should be inclusive and participatory; external evaluators ensure different perspectives are heard
- Use a flexible theory of change (which is revisited and redefined) and a nested approach to assess multiple levels

- Aim to help assess if the policy/initiative is contributing to moving towards its objectives
- Be integrated with policy design and implementation
- Support learning and reflexivity

The herewith proposed contribution aims at providing insights into the evaluation of INNOVATORINNEN, which, in large parts (and as argued in the EuSPRI 2023 contribution), comes up to requirements of TIP-evaluations according to Wise et al. (2022). While demonstrating the evaluation methodology, particular focus will be given to the close collaboration between evaluators and program owners. While ongoing, the program has been strongly responsive to evaluation results. Evaluators and commissioners acted as equal partners in a collaboration that was strongly focused on the content-related development of the program.

In particular, it shall be demonstrated how outcomes of the evaluation have impacted the further development of the INNOVATORINNEN program. For example, since participants of the INNOVATORINNEN Leadership program (a 10-month course for a set of selected participants including coaching for personal development as well as working on an individual professional mission) were subject to surveys and interviews, interim results led to adaptations to the Leadership program, partly even while ongoing. Further immediate outcomes of the evaluation were the parameters to develop the INNOVATORINNEN Club (an element of the program that contains information and networking opportunities for the target group), as well as, most recently, the development of the INNOVATORINNEN Lab. The latter must be considered an immediate response to the interim results of the evaluation, which stressed that the dissemination and (commercial) exploitation of research results were challenging for female researchers and insufficiently supported, promoted, and accompanied in the Austrian research and innovation system. Moreover, it supports the respondents' desire to create true social, ecological, and economic impact and sustainability with their research, since the INNOVATORINNEN Lab will equip its participants with systemic design skills for providing strong impulses within their innovation system.

The contribution will delve into these aspects and highlight the reflective and learning character of the program based on controlled trial and testing (Alber et al., 2021). With a view to the evaluation, the process is characterised by mutual learning and knowledge transfer between evaluators and programowners, which does not only provide a basis for informing and refining the transformation process but also forshaping the discourse of empowerment and visibility of female researchers and innovators from an intersectional perspective. In this light, it can be expected that the analysis will result in a contribution to the current model of TIP evaluations, particularly with a view to mutual learning and knowledge transfer.

REFERENCES:

- Alber, C., Dusl, L., Ecker, B., Pohoryles-Drexl, S. (2021). Erfahrungen und Ergebnisse aus der begleitenden Erhebung zumPilot w-fFORTE Innovatorinnen.URL:https://repository.fteval.at/570/1/w-fFORTE%20Innovatorinnen_Erfahrungen_Pilot_BMDW_FFG_WPZResearch.pdf
- Boni, A., Giachi, S., and Molas-Gallart, J. (2019). Towards a Framework for Transformative Innovation Policy Evaluation.Transformative Innovation Policy Consortium (TIPC) Research Report (April 2019)
- Chataway, J., Daniels, C., Kanger, L., Schot, J., and Steinmueller, E. (2017).'Developing and Enacting TransformativeInnovation Policy'. Paper presented at the 8th International Sustainability TransitionsConference, Gothenburg,Sweden, 18–21 June
- Ghosh B., Kivimaa P., Ramirez M., Schot J., Torrens J. (2021) 'Transformative Outcomes: Assessing and ReorientingExperimentation with Transformative Innovation Policy', Science and Public Policy, 48: 739–18.
- Greussing, E., Schott, M. (2016). Austrian University Female Founders Report 2016.WU Wien, URL:https://www.wu.ac.at/fileadmin/wu/d/cc/gruenden/Report_Female_Founders_16.pdf
- Klapfer, K. & Moser, C. (2022). Arbeitsmarktstatistiken. Ergebnisse der Mikrozensus-Arbeitskräfte-Erhebung und derOffene-Stellen-Erhebung.URL:https://www.statistik.at/fileadmin/publications/Mikrozensus-Arbeitsmarkt-2021.pdf
- Molas-Gallart, J., Boni, A., Schot, J., and Giachi, S. (2020). A Formative Approach to the Evaluation of TransformativeInnovation Policy. Transformative Innovation Policy Consortium (TIPC) Research Report (July 2020).
- Molas-Gallart, J., Boni, A., Giachi, S., and Schot, J. (2021) A Formative Approach to the Evaluation of TransformativeInnovation Policies, Research Evaluation, 2021: 1–12.

- OECD. (2016). Kurzdossier zum weiblichen Unternehmertum.URL:https://www.oecd.org/cfe/leed/Policy%20Brief%20on%20Women%27s%20Entrepreneurship%2DE.pdf
- Patton, M. Q. (2006) 'Evaluation for the Way We Work', Nonprofit Quarterly, 13:28–33.
- Peersman, G. (2015). Impact evaluation, in:https://www.betterevaluation.org/themes/impact_evaluation
- Régent, V. (2023). Evaluating transformative innovation policy instruments with the example of the Austrian programINNOVATORINNEN. Presented at Eu-SPRI Conference, June 2023, Brighton, UK.
- Régent, V., Alber, C., Ecker, B., Pohoryles-Drexel, S. (2023).Erfahrungen und Ergebnisse aus der begleitenden Erhebungdes Programms INNOVATORINNEN, in: Wroblewski et al. (Hg.). Von der Geschlechterpolitik zurdiversitätsorientierten Gleichstellungspolitik. Wien: BMBWF
- Régent, V. & Ecker, B. (2024). INNOVATORINNEN–Begleitforschung des Programms 2022-2023. Final report(forthcoming)
- Schot, J., Boni, A., Ramirez, M., and Alvial-Palavicino, C. (2019). 'Transformative innovation policy and social innovation'.In: Howaldt, J., Kaletka, C., Schröder, A., and Zirngiebl, M. (eds.) Atlas of Social Innovation, ii: A World of NewPractices.Munich: oekom Verlag.
- Wise, E., Eklund, M., Smith, M., Wilson, J. (2022).A participatory approach to tracking system transformation in clustersand innovation ecosystems–Evolving practice in Sweden's Vinnväxt programme, in: Research Evaluation,31(2), 2022, 271-287.https://doi.org/10.1093/reseval/rvac006
- Wisenöcker, A., Resl, B., Szentgyorgyi, R., Borchardt, K., Seidl, A., Fitzgerald, A. (2021). Frauen in der Medizin 2021.Frauen in Führungs-und Entscheidungspositionen in der Medizin. Was fördert und was hindert die Karriereentwicklung? Karl Landsteiner Institut für Human Factors & Human Resources im Gesundheitswesen.URL:https://www.kli-hr.at/wp-content/uploads/2022/03/Frauen-in-der-Medizin-2021_%C2%A9-KLI-1.pdf
- Wroblewski, A. (2022). Leitfaden zur Entwicklung von Gleichstellungsplänen in österreichischen Hochschul-undForschungseinrichtungen. Studie im Auftrag des Bundesministeriums für Bildung, Wissenschaft und Forschungund des Bundesministeriums für Klimaschutz, Umwelt, Energie, Mobilität, Innovation und Technologie.

Jakob Kofler, Jelena Angelis, Elina Griniece, Matthijs Janssen, Peter Kaufmann, Alasdair Reid, Harald Wieser
Austrian Institute for SME Research, Austria
European Future Innovation System Centre ASBL, Belgium
European Future Innovation System Centre ASBL, Belgium
Utrecht University, Netherlands
Austrian Institute for SME Research, Austria
European Future Innovation System Centre ASBL, Belgium
Austrian Institute for SME Research, Austria

HOW ARE THE EU MISSIONS ACHIEVING THEIR GOALS? ATHEORY OF CHANGE ASSESSMENT OF SCALING PROCESSES

ABSTRACT
Although the'renewed' interest for mission-orientated innovation policies (MOIP) is not that new anymore, after over a decade of debate (Mowery et al., 2010; Foray et al., 2012; Mazzucato, 2016), there is still much conceptual confusion about what makes them distinct from other innovation policies with 'transformative' ambitions, in both theory and practice (Hekkert et al., 2021; Edler et al., 2024). As a result of ambiguity over the core underpinnings of MOIP, there are different interpretations of what MOIP is or can be (Janssen et al., 2023).This limits their evaluation, opportunities for learning, and critical reflection regarding their suitability vis-à-vis alternative approaches.
To make assumptions transparent and contribute to a better, processual understanding of how MOIPs are supposed to address societal challenges, this paper examines how processes of scaling are being imagined and enacted in MOIP. Scaling is crucial in connecting research and innovation with the ambition of addressing wider societal challenges (Pfotenhauer et al., 2021). Adopting a'theory of change (ToC) approach (Prinsen & Nijhof, 2015; cf. Molas-Gallart et al., 2021; Haddad & Bergek, 2023), we combine insights from literature, mainly Mazzucato's (2018) and 2019 theorising of MOIP, with an analysis of EU Missions to identify the main mechanisms through which the EU Missions seek to o scale solutions. Mazzucato's views on the design of the EU Missions, which have sparked widespread attention for MOIP elsewhere, provide a skeleton of a generic ToCagainst which policy implementation and results can be compared to examine how MOIP practice diverges from theory.
To understand how the EU Missions are aiming to achieve their goals, we draw on empirical material gathered for recently published external assessment (Reid et al., 2023), comprising documentary evidence, workshops, and semi-structured interviews with senior professionals responsible for the implementation of the missions. Our empirical

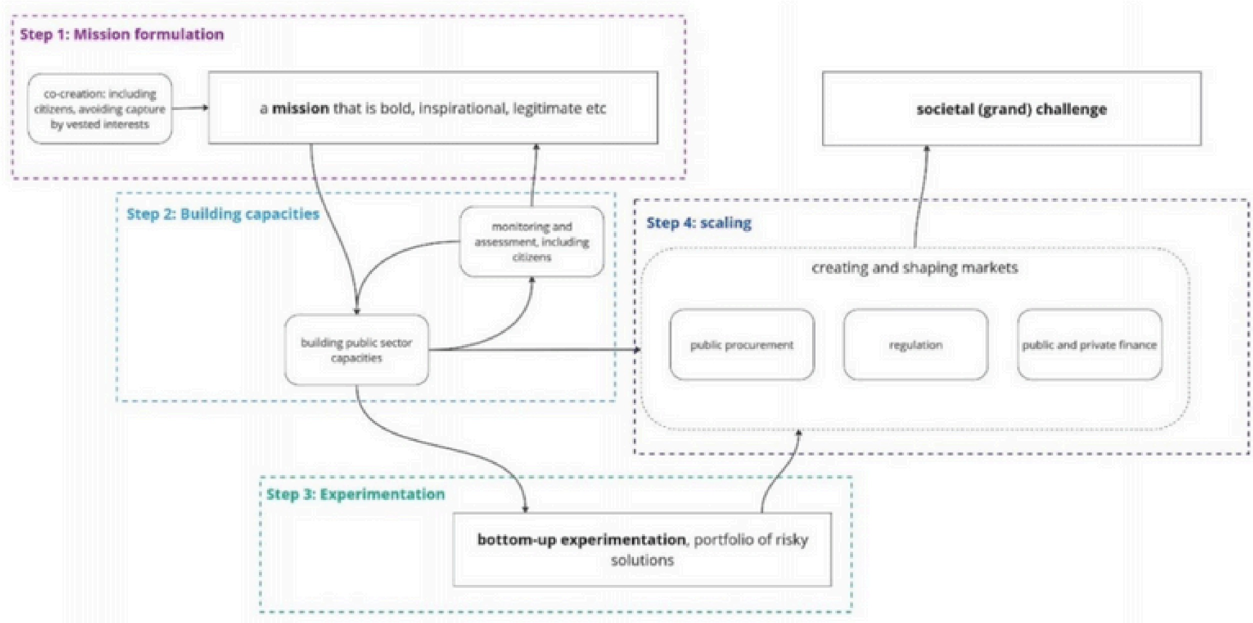
analysis helps to explicate further which change mechanisms (including actions and assumptions) are pursued and how these are similar or different across the five heterogeneous EU missions. The ultimate objective is not to settle debates on what MOIPs are; as empirical realities, there are many different MOIPs. Instead, assuming that TOCs are specific to each MOIP, we aim to highlight practice and variety and thereby demonstrate the use of a TOC approach for assessing (and advancing) MOIPs.

We follow recent publications advocating a ToC approach for evaluating innovation policies with transformative ambitions (Molas-Gallart et al., 2021; Haddad & Bergek, 2023; Rohracher et al., 2023). Such an approach allows for creating a structured overview of the different mechanisms used for achieving a desired impact, with attention for the linkages between what is being undertaken and how this, through a series of causal steps, adds up to impact. Developing a ToC is useful for reflecting on the consistency and completeness of a policy approach; what is missing or overdone?

Combining existing theorising and context-specific stakeholder perspectives, we follow multiple steps in developing the TOC. Taking Mazzucato's reports for the EC (2018; 2019) as an entry point, we first translated her suggestions and recommendations into a graphic ToC and a corresponding template containing a set of assessment criteria for each building block in the ToC. In the second step of our analysis, we develop TOCs for each mission separately, based on the common template, the intervention logics of mission boards, and the interview reports we revisited. In a third step, we follow an iterative process of developing an analytical framework by contrasting the TOCs with MOIPtheorizing in the literature. On this basis, we identify critical gaps and shortcomings as well as communities and differences in the ToCs.

The figure below provides a graphical synthesis of Mazzucato's argument in the form of a TOC. According to Mazzucato (2018), MOIPs in their most rudimentary form consist of two layers to be derived from a societal challenge: the mission (as a top-down element) and mission projects, or a portfolio of projects and experiments (as bottom-up elements). The combination of missions and projects exhibits a push-pull dynamic (cf. Janssen, 2022). Moving beyond this basic structure, ToCneeds to explicate how missions and R&I projects are defined and how these two elements interact to address a societal challenge. In the second report to the EC, Mazzucato (2019) lays out multiple processes that showmissions need to be governed and implemented. We break the TOC down into four building blocks to facilitate the analysis and comparison of EU Missions.

FIGURE 1: TOCOFMOIPS, BASED ONMAZZUCATO (2019) (OWN ELABORATION):



As depicted in Figure 1, we suggest that Mazzucato's recommendations to the EU framed the scaling challenge in terms of market creation and shaping. Three mechanisms stand out: public procurement, regulation, and the mobilisation of public and private finance. As the study is still work in progress, we are unable to show synthesised results at this stage. Our preliminary analysis shows that the intervention logics prepared by mission boards and the ongoing implementation of the missions considered a much wider range of scaling mechanisms, in particular learning across leaders and followers, the translation and embedding of solutions to new local contexts, awareness-raising among stakeholders, and alliance-building.

So far, we can conclude that the EU missions pursue impact in rather distinct ways (e.g., by following an elaborated plan or by acting as merely an open ‘attractor’). Moreover, the associated ToCs go in several ways beyond the mechanisms suggested by Mazzucato, calling into question the emphasis placed on market creation and shaping in the MOIP literature. At the same time, the approaches adopted to scaling seem underdeveloped in several missions. Overall, the EU missions generally rely on rather comprehensive and coherent sets of targeted policy actions, grounded in R&D policy but stretching much further. This could inspire national and regional programs that otherwise might be too ‘narrow’. We plan to conclude by engaging with contemporary critical debates (Brown, 2021; Hekkert, 2023) and providing a nuanced reflection on the suitability of MOIPs for meeting societal challenges.

Hendrik Berghäuser
FraunhoferInstitute for Systems and Innovation Research, Germany

COMBINING EXCELLENCE AND SOCIETAL IMPACT IN RESEARCH ASSESSMENT: DO CHANGE AGENTQUALITIES MAKE A DIFFERENCE?

ABSTRACT

As research increasingly seeks to contribute to solutions for societal problems, with strong emphasis on social engagement (Belcher et al. 2016, Bornmann 2013), research funders are faced with the challenge to design funding programs generating societal impact and research impact. This implicates a need for research assessment processes that take into account both scientific excellence and the potential of societal impact and that are able to deal with tradeoffs between both, especially when diverse actors and researchers outside mainstream circles are addressed (Kraemer-Mbula et al. 2020; Ferretti et al. 2018). This paper provides insights on an ongoing evaluation of a research fellowship program that aims to enable excellent scientists from developing and emerging countries to spend their research period in Germany. 4 The fellowship program wants to recognise the relevance of researchers from these countries in achieving the 17 UN Sustainable Development Goals (SDGs) (United Nations 2015). Thus, classic indicators of scientific excellence such as key publications, scientific impact (h-index), etc. need to be contextualised, given the heterogeneity of researchers from 137 eligible countries and the partly difficult conditions in the science systems in the global south (Tijssen and Winnink 2022). At the same time, the notion of "societal impact" or "contribution to SDGs" appear abstract and elude standardised, indicator-based objective evaluation. Therefore, in addition to scientific merits, the jury also looks for change agent and multiplier characteristics in the applicants (see, for example, Ottaway 1983, Caldwell 2005). This is linked to the assumption that these personal characteristics increase the likelihood that the results of the funded research projects will eventually find their way into societal application in the home countries and thus generate a development-relevant impact.

The focus of this paper is twofold: First, it wants to examine and discuss possibilities and limits when linking two different funding objectives—scientific excellence and societal or developmental impact—as well as potential conflicting goals. Second, the paper seeks to examine whether the funded research fellows actually have more pronounced change agent characteristics than comparable non-funded scientists and in how far these characteristics help translate research findings into actual societal application.

The methodological approach used here is based on a triangulation of different quantitative and qualitative as well as reactive and non-reactive empirical methods. The analysis relies on participant observation of two meetings of the research assessment jury based on an observation guide, five interviews with program officers and representatives of the research assessment jury, an online survey of all funded research fellows from 2010

of the research assessment jury, an online survey of all funded research fellows from 2010 to 2020 within that program, and a control group survey with non-funded scientists from developing and emerging countries with comparable characteristics identified by using bibliometric methods.

As interviews and participant observation have shown, research assessment in this funding program applies a context-sensitive approach in the use of assessment criteria. Not only are indicators of scientific excellence placed in individual circumstances. Also, the societal relevance of the proposed research project and the individual potential of the applicants are evaluated in an individual context. The online survey of the funded fellows (n = 502) confirmed a strong societal orientation and motivation in research and distinct change-agent characteristics among the participants. For example, over 90% of the participants indicated that societal relevance and impact should be a rather important or very important priority in the science system. Over 80% stated that cooperation with non-scientific actors plays an important role in research. In addition, more than two-thirds of the respondents believed that scientists should be actively involved in public debates. Besides measuring personal attitudes, the survey also showed a strong engagement of fellows in transfer and exchange activities. The vast majority of participants engage in exchange with actors from politics, business, practitioners, citizens, NGOs, the media, as well as cultural, educational, and artistic institutions. They are particularly active in the area of science communication (mentioned by over 90% of the respondents), in training and education programs for non-scientists, and in research collaboration with non-scientific partners (75% each). The results from the control group survey were yet not available at the time of submission.

As the case study of this funding program shows, the funding goals of scientific excellence on the one hand and societal impact can go hand in hand, although not all conflicting goals can be resolved in detail. For some, e.g., highly fundamental and theoretical research, the proof of developmental relevance often remains more difficult than for application and transfer-orientated research projects, whose results promise immediate solutions for concrete societal challenges. The key to a successful linkage of scientific excellence and societal impact in research assessment, however, lies in the application of a broad understanding of both conceptual constructs. In this sense, the assessment of scientific excellence should look beyond the use of classical parameters related to publication output and performance and include assessments of a personal research profile and of the scientific potential of the researcher as such. At the same time, a broader understanding of developmental relevance and societal impact also broadens the view of research, whose societal impact cannot be clearly assessed now, as is often the case in fundamental research, for example. This broader understanding of impact also prevents the risk of placing an (overly) strong focus on application-orientated research only in developing and emerging countries and acknowledges instead the importance of fundamental research for the development of research systems in these countries. In addition, considering change agent characteristics of applicants seems to be a promising approach in research assessment. The online survey among funded fellows showed a very strong commitment to transferring research into social practice. The participants also maintained a close exchange, especially with those societal actors for whom their research is particularly relevant.

REEFERENCES:

The paper is based on an evaluation of the Georg Forster Research Fellowship providedby the Alexander vonHumboldt-Foundation in Germany. For more information about the funding program, see:https://www.humboldt-foundation.de/en/apply/sponsorship-programmes/georg-forster-research-fellowship.

- Bornman,L. (2013): What is societal impact of research and how can it be assessed? A literaturesurvey. Journal of the American Society for Information Science and Technology 64(2), 217-233.
- Belcher, B.; Rasmussen, K.; Kemshaw, M.; Zornes, D. (2016): Defining andAssessing Research Quality in aTransdisciplinary Context. Research Evaluation 25, 1–17.
- Caldwell, R. (2005): Agency and Change. Rethinking Change Agency in Organizations. Routledge. London.
- Ferretti, F.; Pereira, A.; Vertesy, D.; Hardeman, S. (2018): Research excellence indicators: time toreimagine the 'making of'? Science and Public Policy 45 (5), 731–741.
- Kraemer-Mbula, E. et al. (Eds.) (2020): Transforming Research Excellence. New Ideas From the GlobalSouth. Cape Town, South Africa.
- Ottaway, R.D.(1983): The Change Agent: A Taxanomy in Relation to the Change Process. In: HumanRelations 36(4), 361-392.
- Tijssen, R.; Winnink, J. (2022): Global and Local Research Excellence in Africa: New Perspectives onPerformance Assessment and Funding. In: Science Technology and Society 27(3), 368-387.
- United Nations (2015): Transforming our world. The 2030 Agenda for Sustainable Development.A/RES/70/1:https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf.

Michael Ornetzeder, Steffen Bettin, Jakob Kofler, Michael Wedler, Harald Wieser
Austrian Academy of Sciences, Austria
Austrian Academy of Sciences, Austria
Austrian Institute for SME Research, Austria
B.A.U.M. Consult GmbH, Germany
Austrian Institute for SME Research, Austria

A READINESS ASSESSMENT FRAMEWORK FOR URBAN SYSTEM INNOVATION

ABSTRACT

Making cities more sustainable requires ground-breaking solutions that are systematically integrated into existing structures (Castán Broto and Bulkeley 2013). Consequently, research funding must focus not only on the development of new technologies but increasingly on overall systems that are working well in the real world. However, the shift from technology-focused solutions to more holistic approaches presents several challenges. These include how to adequately describe the intended innovations and how to produce knowledge about their state of development.

Motivation and context to study the readiness of urban system innovation In this paper are the various activities in the field of sustainable urban development (European Union 2011), especially in the context of the Austrian Smart Cities Initiative (Klima-und Energiefonds 2022) and the national mission-orientated research and innovation (R&I) initiative "climate-neutral ties." European cities face a number of challenges and demands. Innovation in the urban context plays a key role. In this debate, innovation is seen not only as a normative concept, as Godin and Gaglio (2019) rightly point out, but also as a programmatic one, as there is a strong expectation that innovation will be able to solve these pressing problems. As a result, there is a shift in perspective from assessing the readiness of technologies to assessing the readiness of urban innovations.

Readiness assessment is a well-established field of research and a widely used consulting practice (Olechowski, Eppinger et al. 2015). Initially, there was a strong emphasis on evaluating the readiness of technologies deemed crucial to the success of innovative solutions (Mankins 1995, Mankins 2009). However, early criticism of this technology bias prompted proposals to broaden the scope of assessment to encompass a wider perspective (Tao, Probert et al. 2010, EARTO 2014). The diversity of development contexts has resulted in the development of specific readiness assessment frameworks for distinct contexts (Vik, Melås et al. 2021, Holden 2022, Sprenkeling, Geerdink et al. 2022). Although much has been done in this area in recent years, there is still no approach that is capable of adequately assessing the readiness of system innovations and identifying development

options. In response to this situation, the framework presented in this paper was developed.

The framework presented in this paper was carried out in 2023 on behalf of and in close coordination with our clients, the Austrian Climate and Energy Fund and the Federal Ministry of Climate Action, Environment, Energy, Mobility, Innovation, and Technology. The aim of the project was to develop a practical tool for assessing the readiness of urban system innovations for monitoring and providing strategic advice to national R&D programs.

Empirically, the findings are based on several meetings with the clients, interviews with stakeholders, potential users, and international experts. In addition, preliminary versions of the framework were represented, discussed, and further developed in three workshops with different target groups. Four one-hour interviews were conducted with representatives of ministries (1) and intermediary organisations (3) to assess national needs and potential use cases for readiness assessments. Two online interviews were conducted with readiness assessment experts to learn more about the development of existing frameworks and experiences with their application.

The first workshop was also held online and was used to test the framework for the first time with the developers of urban system innovations. The test focused on the plausibility of the selected dimensions and the implementation of the readiness assessment based on predefined definitions. In the second workshop, the framework was presented to five representatives of intermediary organisations that mediate between research and funding and provide strategic and operational support to the ministry, the fund, and the agency. The focus was placed on the applicability of the framework. The third and final workshop was attended by 12 representatives from different units of the Federal Ministry's innovation and technology department and other national research funding organisations. The workshop focused on the future role of the readiness assessment in the context of national research programs (with a focus on climate-neutral cities) and what further steps would be beneficial for its implementation.

The framework In order to provide a framework for assessing the readiness of USIs, we identified the following seven dimensions as constitutive for systemic innovations: (1) system architecture, (2) technology, (3) production, (4) infrastructure, (5) operation, (6) usage, and (7) rules (see Table 1). These dimensions were derived from the literature review and the study of empirical examples of USIs (literature and workshop) and were designed to be as complete as possible in describing a USI and how it operates in the real world. We have also ensured that the dimensions relate to the innovation itself and not to its environment (with the exception of necessary interfaces). As the aim is to assess the readiness of the innovation, only those factors that can be directly influenced by the developers, research funders, and other social actors involved are relevant in our context. This clearly distinguishes our approach from existing assessment frameworks, which typically include an assessment of the social environment (see Holden 2022, Vik et al. 2021), meaning that innovations can only be considered ready in some dimensions at a very late stage (after their actual market success and wider dissemination). In defining the dimensions, it was also important to keep them as clearly separated as possible and to avoid overlaps. The dimensions include upstream activities (production, planning) as well as key components for the realisation of the targeted functions.

Table 1: THEREADINESSFRAMEWORK FOR URBAN SYSTEM INNOVATIONS

		Dimensions of Urban System Innovations						
Development Phase	**Readiness Level**	Technology	Production	Infrastructure	Usage	Operation	Rules	System Architecture
Concept Development	**1**							
	2							
	3							
Experimentation	**4**							
	5							
	6							
Implementation	**7**							
	8							
	9							

Source: own illustration

To evaluate the readiness of an USI, the readiness of each of the seven dimensions needs to be assessed.For this purpose, we have defined a scale with nine generic levels, following the established definitions of TRL scales (Mankins 2009). The breakdown into a total of nine maturity levels has proved useful in practice (Olechowski, Eppinger et al. 2015). A lower gradation is associated with a significant loss of information, while a more detailed gradation increases the measurement effort without a corresponding improvement in information content. Our scale also follows a 'three-by-three' logic, where the three overarching areas of 'concept development', 'experimentation', and 'implementation are each subdivided into three finer levels. This is to emphasise that development processes consist of fundamentally different phases with different emphases and activities, which are also reflected in different funding models and support services (Kline and Rosenberg

2010). To make the framework more user-friendly, we have developed an individual scale for each of the seven dimensions based on the generic definitions (63 definitions in total). In addition, we have derived questions from each of these definitions to further support the assessment process of USIs.

We have introduced a new framework for assessing the readiness of Urban System Innovations (USIs). USIs are crucial for cities to become more sustainable and climate-friendly. Assessing the readiness of such systemic innovations can make an important contribution to better governance of ongoing innovation activities. The proposal presented here is a contribution to the development of readiness assessment models that go beyond the narrow focus on the role of technologies (EARTO 2014, Vik, Melås et al. 2021, Sprenkeling, Geerdink et al. 2022). The multidimensional framework proposed here helps to make research needs visible where technological maturity is high. In so doing, the framework can lend legitimacy to decisions to fund R&I projects that address social innovations or the architecture of an USI.

However, to overcome the structural bias towards technological solutions in research funding, the framework needs to be implemented at several points in the R&I funding system: 1) in funding proposals, 2) in project evaluation criteria, 3) in the preparation of calls for proposals, 4) in the public procurement of innovative solutions, 5) in the presentation and communication of innovations to investors, and 6) in monitoring and evaluation. Multiple implementation challenges need to be overcome, related to the mediation between the assessment of the level of R&I projects and the level of USI, the lack of non-technical knowledge among reviewers, and EU funding rules. In the presentation, we discuss these implementation challenges and present ideas on how the presented assessment framework could be integrated into existing funding structures.

Dr. Elizabeth Gadd
INORMS Research Evaluation Group/Loughborough University, United Kingdom

THE 'WICKED' PROBLEM OF GLOBAL UNIVERSITY RANKINGS AND HOW TO FIX IT

ABSTRACT
A less well-discussed element of the move towards more responsible research assessment (RRA) is the question as to how we might better provide university-level assessment and specifically mitigate the cascading, negative impacts of global university rankings on department, research group, and researcher-level assessment. The poor assessment mechanisms used by the global rankings have long been understood by the academic community (Hazelkorn, 2013), but the appetite and perceived challenge of the rankings has been limited. The expression ‚rankings are here to stay‘ has been pervasive, and the overarching mood is one of helplessness.

To raise awareness of and ultimately seek to address these challenges, the INORMS Research Evaluation Group developed two initiatives. Firstly, in 2020, they developed some‚Ranker Ratings'to highlight the discrepancy between community-developed expectations of responsible university-level assessment and the actual performance of global university rankings (Gadd, 2020). More recently, they introduced the More Than Our Rank initiative (INORMS REG, 2023) by which universities can highlight how much more they have to offer the world than is captured by the global university rankings.

Perhaps in response to this work, the problem of rankings has started to be acknowledged in formal RRA policy positions as a significant hindrance to the implementation of RRA. ‚Rethinking the Rankings‘ was one of ten recommendations of the recent ‚Harnessing the Metric Tide‘ Review of indicators, infrastructures & priorities for UK responsible research assessment (Curry, Gadd & Wilsdon, 2022). Whilst‚avoiding the use of university rankings in researcher assessment‘ is the 4th core principle of the CoARA Agreement on Reforming Research Assessment (CoARA, 2022).

Individual HEIs are starting to respond to the limited and misleading view of institutional quality offered by the rankings relative numbers via two main mechanisms. Firstly, global institutions in countries such as China, India, South Korea, the US, South Africa, the Netherlands, and Switzerland are boycotting various rankings by refusing to submit data (see, e.g., Jung & Sharma, 2023; Niazi & Sharma, 2023)

Secondly, an increasing number are joining the More Than Our Rank initiative. This allows institutions to describe in a qualitative way how much more they have to offer the world than is captured in the global university rankings. The CWTS Leiden Ranking has started to identify More Than Our Rank Signatories on their rankings to provide a qualitative

complement to their quantitative data.

Alongside these practical interventions, in 2023, three national and international collectives issued white papers dedicated to making recommendations to a range of stakeholders in the hope of engendering more critical engagement with the global university rankings.

An expert group representing the Universities of the Netherlands (UNL, 2023) produced a recommendation paper entitled ‚Ranking the University, which made 14 recommendations for individual, national, and international institutions.

An international expert collective headed up by the United Nations thinktank, the UN University International Institute for Global Health (2023a, b), issued both a white paper‚Interrogating the Coloniality and Biases of Global University Rankings‘and a subsequent statement. The latter made 12 recommendations to a wide range of stakeholders to help tackle the coloniality and biases of global rankings.

Finally, the European University Association (2023), after a plenary session at their annual conference dedicated to questioning the value of the global university rankings, developed a briefing to member institutions. This outlined tenkey considerations for the use of rankings by higher education institutions, including eight recommendations for action.

The fear for institutions reliant on international student recruitment is that any critical engagement with the rankings could have financial and reputational percussions. The number of actors required to have any significant impact on the rankings status is considerable, and the most powerful actors (e.g., the rankings themselves and the highly ranked) have no real incentive to make change. All considered, this issue shares many of the features of a wicked problem (Rittel & Webber, 1973). As such, it is difficult to confidently identify the root of the problem, the ‚right‘ solution, a mechanism for testing that solution, and to know when the ‚problem‘ has been fixed.

It is proposed that the core agents and actions contributing to the maintenance of global university ranking areas are outlined in Table 1.

Sta ge	Agent	Action
I	Universities & academics	Supply data to rankings and respond to reputation surveys
2	Ranking agencies	Judge universities using inappropriate mechanisms
3	Ranking agencies (& highly ranked)	Promote rankings as 'truth'
4	Students, governments, academics, funders, institutions	Believe the rankings are accurate & make decisions based on them
5	Universities & governments	Try to do better in the rankings

6	Ranking agencies	Sell services to help universities and government agencies do better in the rankings
7	Universities	Promote their rank (lending legitimacy to rankings)

By mapping the recommendations of existing RRA policies as they relate to rankings, the three whitepapers, and interventions such as More Than Our Rank, to the different stages of the ranking process, we can identify at which stage their initiators consider it might be possible to address this wicked problem, and therefore which groups are seen as having the agency to make these changes.

This session will explore with the audience the extent to which the proposed set of steps, agents, and actions involved in maintaining the current rankings status quo match their experience. An examination of the mapping evidence will support a discussion as to the likely success of various interventions aimed at disrupting the status quo and ultimately mitigating some of the negative impacts of global university rankings.

REFERENCES:

- CoARA (2022).Agreement on Reforming Research Assessment.URL:https://coara.eu/app/uploads/2022/09/2022_07_19_rra_agreement_final.pdf
- Curry, S., Gadd, E., & Wilsdon, J. (2022)Harnessing the Metric Tide:Review of indicators, infrastructures& priorities for UK responsible research assessmentURL:https://rori.figshare.com/articles/report/Harnessing_the_Metric_Tide/21701624
- European University Association (2023).Key Considerations for the Use of Rankings.URL:https://eua.eu/downloads/publications/key%20considerations%20for%20the%20use%20of%20rankings%20by%20heis.pdf
- Gadd, E.(2020). University rankings need a rethink.Nature587, 523URL:https://doi.org/10.1038/d41586-020-03312-2
- Hazelkorn, E.(2014)Reflections on adecade ofglobalrankings: what we've learned and outstandingissues.EuropeanJournal ofEducation49, 1: 12-28.
- INORMS REG (2023).More Than Our Rank.URL:https://inorms.net/more-than-our-rank/
- Jung, U. & Sharma, Y. (2023) Korean universities unite against QS ranking changes.University WorldNews(4 July 2023). URL:https://www.universityworldnews.com/post.php?story=20230704195008557

- Niazi, S. & Sharma, Y. (2023) IITs lead in national rankings but continue global boycott. UniversityWorld News (8 June, 2023)https://www.universityworldnews.com/post.php?story=20230608132444887
- Rittel, H.W.J., & Webber, M.M. (1973) Dilemmas in a general theory of planning.Policy Sci4, 155–169.URL:https://doi.org/10.1007/BF01405730
- UNL (2023). Ranking the University. URL: https://www.universiteitenvannederland.nl/en/ranking-the-university
- UN University International Institute for Global HealthInternational Expert Group(2023)WorldClass Universities? Interrogating the Coloniality and Biases of Global University Rankings. URL:https://collections.unu.edu/view/UNU:9082
- UN University International Institute for Global Health International Expert Group (2023) Statementon Global University Rankings. URL:https://collections.unu.edu/eserv/UNU:9299/Statement-on-Global-University-Rankings.pdf

Gabriel Falcini dos Santos, Sergio Luiz Monteiro Salles-Filho, Ana Carolina Spatti
University of Campinas, Brazil

CONNECTING UNIVERSITY, RESEARCH ORGANISATIONS AND INDUSTRY

ABSTRACT

Innovation is receiving growing attention from policymakers as an important element in addressing contemporary challenges (OECD & Eurostat, 2018). Besides promoting economic growth, innovation can help tackle social progress, gender inequality, and environmental performance, but for that, studies point to the necessity of overcoming an outdated Research and Development (R&D) model for innovation. It becomes mandatory to switch from a profit-driven model to reach a new collaborative model with the engagement of diverse stakeholders (Elisa Arond et al., 2010; Etzkowitz & Leydesdorff, 1995; Galvao et al., 2019). Furthermore, a transformative model of R&D and innovation must take into account the differences between countries in the Global South and North. While the entire world faces social and climate issues, developing countries often need to deal with challenges that developed countries have already overcome (Fu et al., 2011; Lundvall et al., 2009; Mohamed et al., 2022). 4 Within this context, Brazil faces historical challenges in leveraging private R&D investments. Despite having comparable public expenditure in R&D to several developed countries, Brazil grapples with significantly lower private investments. The government's longstanding efforts, including tax incentives and economic subsidies, have not been sufficient (Pacheco & Corder, 2010; Oliveira, 2011; De Negri, 2017; MCTI, 2021). In 2013, the Brazilian government established the Brazilian Company of Research and Industrial Innovation (Embrapii), a private non-profit organisation created to foster collaborative research between Research Organisations (ROs) and Industry. This unique model in the country (Castro et al., 2017; Salles-Filho et al., 2021) can present itself as a model for overcoming this historical scenario of stagnation. Embrapii accredits existing ROs in Brazil, such as laboratories and university departments affiliated with higher education institutions, whether public or private, and nonprofit research institutes. The accredited institutes are called Embrapii Units (UEs), and after accreditation these UEs seek partnerships with companies to develop pre-competitive technological innovations aligned with the interests of the companies. Financial resources are released only after contracts between the ROs and companies are established To understand the outcomes of the Embrapii model for leveraging Brazilian investments in innovation, we conducted an Impact Evaluation adhering to the highest evaluation quality standards. According to Gertler et al. (2016), evaluations take place periodically to answer questions related to the design, implementation and results of programmes already completed or in progress. According to the same author, Impact Evaluation is a

specific type of evaluation that necessarily works with cause and effect. In it, the question aims to attribute a causal effect of a programme or intervention to a specific outcome of interest.

To operationalise our research, we had access to exclusive data provided by Embrapii itself: the study is based on a sample of 840 projects completed between 2015 and 2021 by 47 UEs and 565 companies. A questionnaire was submitted to the agents involved, and answers were collected from 179 companies (28% of the sample) and 43 UEs (92% of the sample). Additionally, 6 interviews were conducted with agents from companies and 4 interviews with representatives of UEs. We also conducted an extensive literature review about Embrapii and similar models. Only 21 studies were found.

Through data analysis and with the support of input-output analysis and a cost-benefit analysis, we seek to assess the impacts of Embrapii with the following indicators: generated innovations, with emphasis on the percentage of innovations that reduced environmental impact; return on investment; learned competencies; and propensity to invest in R&D before and after the project.

Our results indicate that 68% of projects conducted under the Embrapii model generated at least one innovation, of which one-third led to reduction in environmental impacts. Furthermore, 52% of the companies said that Embrapii contributed from 60 to 100% to achieving innovation; 32% said they would not have undertaken the project without Embrapii, and 30% said they would have done it anyway - figures that respectively indicate signs of crowding in and crowding out of the public investment made. Our study also indicates that for every 200 thousand dollars invested by Embrapii, 450 thousand dollars are added to the Brazilian GDP and 33 jobs are created; and for every dollar invested by companies, 2.7 dollars are returned to them.

Based on the results obtained, we conclude that Embrapii has a positive impact on innovation generation in partner companies, and in part of them, Embrapii leverages private investment in innovation. Our data points to a greater impact of Embrapii on small businesses: in the 2022 the Brazilian Innovation Survey (PINTEC), only 63% of companies of this size achieved innovations, while in our study, 85% did so. As for large companies, in PINTEC, 77% achieved innovations, while among those supported by Embrapii, this value is 62%.

The Impact Evaluation is limited by the need to disregard the influence of external factors to establish a causal relation. Therefore, we can suggest causality in certain aspects, with reservations. Furthermore, information is limited, as 72% of the surveyed companies did not respond to the questionnaire. Nevertheless, we hope to contribute with an original work on a model with particular characteristics specific to the Brazilian reality, which may inspire policy makers from other countries, especially in the Global South. It is worth noting that none of the 21 studies found about the model conducted such a comprehensive evaluation with data directly obtained from Embrapii, highlighting the importance and uniqueness of our study.

REFERENCES:

- Castro, F. P. de, Campos, G. T. de, & Gilaberte, T. P. (2017). A EMBRAPII como perspectiva à inovação ["EMBRAPII as a perspective on innovation."]. Cadernos de Prospecção [“Prospective notebooks”], 10(2), 164–176.
- de Negri, F. (2017). Por Uma Nova Geração de Políticas de Inovação no Brasil ["Toward a New Generation of Innovation Policies in Brazil."]. In Políticas de Apoio à Inovação Tecnológica no Brasil: Avanços recentes, limitações e propostas de ações ["Policies for Technological Innovation Support in Brazil: Recent Advances, Limitations, and Proposed Actions."] (pp. 25–46).
- Elisa Arond, A. Ely, Martin Bell, M. Leach, Ian Scoones, & Andy Stirling. (2010). Innovation, Sustainability, Development: A New Manifesto. STEPS Centre. https://steps-centre.org/publication/innovationsustainability-development-a-new-manifesto/
- Etzkowitz, H., & Leydesdorff, L. (1995). The Triple Helix—University-Industry-Government Relations: A Laboratory for Knowledge-Based Economic Development. 14(1), 14–19.
- Fu, X., Pietrobelli, C., & Soete, L. (2011). The Role of Foreign Technology and Indigenous Innovation in the Emerging Economies: Technological Change and Catching-up. World Development, 39(7), 1204– 1212. https://doi.org/10.1016/j.worlddev.2010.05.009
- Galvao, A., Mascarenhas, C., Marques, C., Ferreira, J., & Ratten, V. (2019). Triple helix and its evolution: A systematic literature review. Journal of Science and Technology Policy Management, 10(3), 812–833. https://doi.org/10.1108/JSTPM-10-2018-0103
- Gertler, Paul J, Sebastian Martinez, Patrick Premand, Laura B Rawlings, and Christel M J Vermeersch. (2016). Impact Evaluation in Practice. 2nd ed. World Bank Training. Washington, DC: Inter-American Development Bank and World Bank. http://hdl.handle.net/10986/25030.
- Lundvall, B.-Å., Vang, J., & Joseph, K. J. (2009). Innovation System Research and Developing Countries. Em B.-Å. Lundvall, K. J. Joseph, C. Chaminade, & J. Vang (Orgs.), Handbook of Innovation Systems and Developing Countries. Edward Elgar Publishing. https://doi.org/10.4337/9781849803427.00006

- Mohamed, M. M. A., Liu, P., & Nie, G. (2022). Causality between Technological Innovation and Economic Growth: Evidence from the Economies of Developing Countries. Sustainability, 14(6), 3586. https://doi.org/10.3390/su14063586
- OECD & Eurostat. (2018). Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation, 4th Edition. OECD. https://doi.org/10.1787/9789264304604-en
- Oliveira, M. B. de. (2011). O inovacionismo em questão [“Innovation-ism in Question”]. Scientiae Studia, 9(3), 669–675. https://doi.org/10.1590/S1678-31662011000300011
- Ministério da Ciência, Tecnologia e Inovações [“Ministry of Science, Technology, and Innovations”] - MCTI (Ed.). (2021). Indicadores Nacionais de Ciência, Tecnologia e Inovação [“National Indicators for Science, Technology, and Innovation”].
- Pacheco, C. A., & Corder, S. (2010). Mapeamento institucional e de medidas de política com impacto sobre a inovação produtiva e a diversificação das exportações ["Institutional Mapping and Policy Measures Impacting Productive Innovation and Export Diversification"]. p 99.
- PINTEC - Pesquisa de inovação [“Innovation survey”]. (2022) Instituto Brasileiro de Geografia e Estatística [“Brazilian Institute of Geography and Statistics”] - IBGE. Rio de Janeiro.
- Salles-Filho, S., Bin, A., Bonilla, K., & Colugnati, F. A. B. (2021). Effectiveness by Design: Overcoming Orientation and Transaction Related Barriers in Research-Industry Linkages. Revista de Administração Contemporânea [“Contemporary Administration Journal”], 25(5), 22.

Christina Schuh, Sonja Fringes, Daniel Schwertfeger
Hochschule des Bundes für öffentliche Verwaltung (HS Bund), Germany

EVALUATING "LEARNING AND EXPERIMENTAL SPACES" - WHEN A TRADITIONAL APPROACH REACHES ITS LIMITS

ABSTRACT

The paper reflects on the experiences from the perspective as evaluators for the overall evaluation of the funding guideline of the Federal Ministry of Labour and Social Affairs: "Sustainable companies and administrations in digital change". Through learning and experimentation spaces (LES), small and medium-sized enterprises are supported in developing innovative, tailor-made and consensual solutions for employees and companies in the digital transformation that are supported by social partnerships. The learning and experimentation spaces should be of a fundamentally exemplary nature and transfer to further innovative solutions. Funding was provided for 17 LES in the first funding round (starting 2018) and a further 11 LES with a focus on the use of artificial intelligence (AI) in the second funding round (starting 2020). The duration of each funded project was about 3 years on average. In the call for proposals for the funding guideline published in the Federal Gazette (BAnz AT 14.08.2017 B2 and BAnz AT 20.01.2021 B2), it was stipulated that "All projects must be evaluated during the entire project term by a body independent of the project recipient that is commissioned by the funding recipient." (funding guideline point 4.2) and "The BMAS [Federal Ministry] will subject all funded measures to an external overall evaluation. This serves to check the extent to which the innovation transfer of the funded learning and experimentation spaces to other companies is possible or successful ..." (funding guideline point 6.8). The Federal University of Applied Administrative Sciences (HS Bund) was commissioned with the overall evaluation, that will be completed in summer 2024.
We used a Multi-Method-Design (see Figure 1):

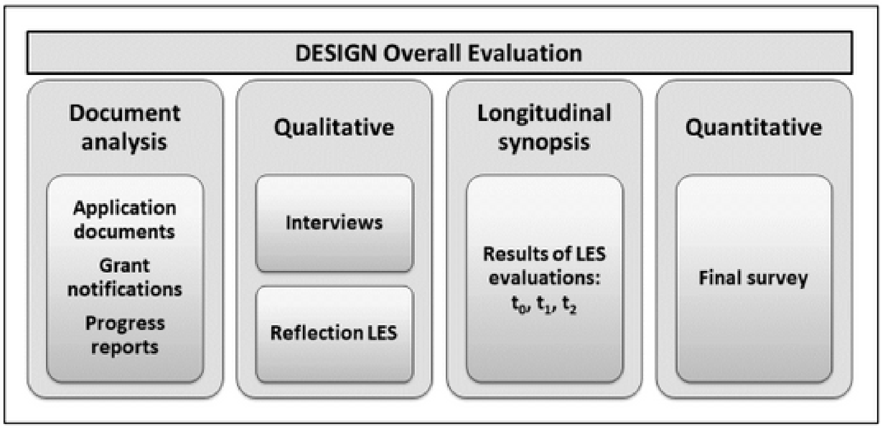


FIGURE 1: DESIGN OVERALL EVALUATION LES

Quite fundamental for funded projects we took the relevant documents relating to the funding process into account. This includes application documents, grant notifications of the Federal Ministry as well as progress reports. According to the mission of the overall evaluation we worked closely with the singular external evaluations by including for example their evaluation concepts as well.
Furthermore, as a qualitative methodological approach we conducted interviews with representatives from all relevant stakeholders related the general funding guideline. Various aspects of the funding guideline were addressed, e. g. its main objectives, factors for success of a LES and the interplay of the stakeholders. Considering a LES as funding instrument, we then asked persons responsible for a funded project to reflect on it using three general questions on its characteristics, its strengths and weaknesses as well as potentials to improve LES as funding instruments. The respective formative evaluation approaches for each LES led to singular survey periods. The challenge for the overall evaluation therefore had to find a way to conduct the results, by synthesizing the data from first 17 then 11 projects in a longitudinal synopsis. From the perspective of data analyses the synopsis stands between the qualitative as well as the quantitative approach. About the latter we mainly conducted a standardized final survey asking project management and external evaluation aiming the main interests of the overall evaluation in terms of planning a LESProject, implementing it and focussing on the transfer of its findings and testing hypothesis that we developed during the formative phase of the evaluation.
The discussion and classification of our experiences as overall evaluators will be based on the following five factors: Methods, role of evaluators, funding logic, context and intervention.

While traditional evaluation approaches are based on a linear funding logic and indicators and methods for measuring outcomes and impact are defined at the beginning of the evaluation project, methods in the developmental evaluation approach1 are adaptive and changeable. We started our overall evaluation with a set of indicators developed by the Federal Institute for Occupational Safety and Health (BAuA). This set of indicators has also been suggested as a methodological frame of reference to the individual external evaluations. However, during the course of the projects the initial concept was increasingly scrutinized and flexibly adapted to the requirements of the overall evaluation in the course of the project

In classic evaluation, evaluators monitor the intervention and compliance with scientific standards, they are independent and often external. In developmental evaluation, the evaluator is seen more as a "critical friend" , part of the team that carries out the intervention. From this perspective methodological approaches and measurements are adapted in recognition of dynamic developments during the course of the project. We received feedback on both role approaches from members of the individual projects.

The funding logic in classic evaluations is seen as linear. Thus, why linear causal effects are sought or causal logic models are set up. Counterfactual comparisons are also made. The developmental evaluation approach tends to look for patterns to prove causality. Several solutions can be tried out and the funding process allows for iterative feedback loops. A LES is characterized by the fact that the learning space offers the opportunity to try out solutions and that the best solution has not yet been determined at the start of the project. The focus is on the learning experience and less on a profitable product.

Traditional evaluation approaches generally attempt to examine a stable situation and control uncertainty. Developmental evaluation focuses on complex and dynamic environments where uncertainty is allowed[1] . In learning and experimental spaces, the aim is to try out solutions and failure of the intervention is also permitted. However, the question arises as to what extend the permission to try out solutions for digital transformation is perceived by the projects. A further question is whether the administrative framework conditions are designed in such a way that the openness for results of various kinds can actually be realized. Especially when it comes to results in the sense of transferable learning experiences without a corresponding tangible product.

While classic evaluation approaches assume a stable intervention, the intervention in the developmental evaluation approach is changeable. In learning and experimental spaces interventions, solutions for digital transformation should be adapted and can be changed.

Learning and experimental spaces (LES) are currently a widely used project funding instrument at municipal, state and federal level in Germany. Their evaluation requires a new perspective on the funding process, the role of the stakeholders and applied methods. For this, the developmental evaluation approach offers important insights.

REFERENCES:

- Patton, M. Q. (2015). State of the Art and Practice of Developmental Evaluation: Answers to common and recurring questions. In: Patton, M. Q., McKegg, K. & Wehipeihana, N. (Ed.): Developmental Evaluation Exemplars. Principles in Practice. The Guilford Press, pp. 1-24.
- Sager, F., Hadorn, S., Balthasar, A. & Mavrot, C. (2021). Politikevaluation. Eine Einführung. Wiesbaden: Springer Fachmedien. DOI: 10.1007/978-3-658-32490-2.

Renée van Dis, Mireille Matt, Allison Loconto
INRAE-LISIS, France

EVALUATING "LEARNING AND EXPERIMENTAL SPACES" - WHEN A TRADITIONAL APPROACH REACHES ITS LIMITS

ABSTRACT

In mission-oriented contexts, researchers are funded to contribute to resolving challenges in society. Going beyond research efforts that are oriented towards economic growth and blue-sky objectives (Joly, 2018), it asks researchers to rethink associations between their science and society and take on responsibility in this regard. We find two widely acknowledged paradigms in academic literature that help understanding alternative science-society associations: 'Science in' (e.g. Latour & Woolgar, 1979; Giddens, 1984; Jasanoff, 2004) and 'Science for' Society (e.g. Schot & Steinmueller, 2016; Kuhlmann & Rip, 2018). Even though this literature stresses the need for systemic changes, it emphasises the role of researchers -and less so the active role of other actors- as well as the difficulties in directing excellent science towards societal goals. Building further upon these two strands, we argue for a third understanding namely: Science with Society. Despite its mention through the EU's SWaFS call, this formulation of the science-society relationship has not been widely problematised in literature. We can learn from previous literature that the achievement of a societal mission is not the direct outcome of a research programme. Instead, contributing to a societal mission is a complex and uncertain process that requires the translation of research results into expected societal transformations (Robinson & Mazzucato, 2019). This process needs - beyond additional time, actors and resources – visions of change in society.

In this paper, we question how researchers could be guided in the process to take on responsibility for contributing to a societal mission, by rethinking the associations between their science and society. We do so through the empirical case of ASIRPA Real-Time (RT), a real-time impact assessment approach implemented in the French Priority Research Programme 'Growing and Protecting crops Differently' (PPR-CPA), with the ambitious mission of eradicating pesticides by 2040. We worked with the researchers of the ten funded research projects who are encouraged to think about their contribution to the constitution of a pesticide-free society as they study alternative solutions. As the conceptual foundation of ASIRPA RT is based on the Sociology of Translation (Callon, 1986), we explore the researchers' responsibilisation as the anticipation of a process of translation. After we highlight the literature we reviewed and the methodological and conceptual implications, we discuss our main results and the conclusions we draw from this.

The notion of responsibility in Research and Innovation

Since the end of WWII, the notion of responsibility in research has been of scholarly interest, starting first with notions of research ethics and moral responsibility. This rise is seen as a consequence of the use of the atomic bomb at the end of WWII, and the questioning of research ethics and moral responsibility in this regard (Strydom, 1999; Douglas, 2003; Temri, 2008). Over time, the understanding of responsibility evolved in line with changing societal challenges such as increasing negative impacts of humans on the planet, the arrival of 'new social movements' and related public opinions, and rising discussions around risk societies (Strydom, 1999). However, what it means to act responsibly can mean different things in different contexts (Vincent, 2010). We can argue that researchers have role responsibilities in the 'duties' they hold in the production of scientific knowledge. In mission-oriented contexts, researchers also have a responsibility with regards to the outcomes of their research, and the impact it has in society. Here we follow Douglas (2003), who argues for a moral responsibility of researchers in their professional role in science. In addition, Arnaldi & Gorgoni (2016) emphasise an 'active' understanding of 'prospective' responsibility that the authors define as the "actors' capacity of self-commitment towards some goals which are not mandated by rules (legal or of other sort)" (p.6). The authors perceive responsibilisation as a 'governance strategy', and that strategies for responsibilisation aim at the ex-ante assumption of responsibility by actors in their commitment to societal goals (Dorbeck-Jung & Shelley-Egan, 2013; Arnaldi & Gorgoni, 2016).

Hence, beyond ethical and moral responsibility, we ask researchers today to contribute to ambitious societal missions with their research projects. Researchers are required to rethink the complex associations between their research and the future society that will embed the alternative solutions. This means that, beyond considering the negative consequences of research and technology on society, researchers are demanded to actively contribute to complex transformative societal change. In such mission-oriented contexts, we expect researchers to assume prospective forms of role responsibilities in contributing to societal impacts. However, how could researchers be guided in the direction of societal missions? And, what does it mean for researchers to take on this type of broader responsibility? To respond to this knowledge gap, we illustrate the process of how integrating formative evaluation in mission-oriented research projects enables researchers to take on responsibility as it guides them in anticipating a process of translation.

The process of Translation: ASIRPA Real-Time as a way to responsibilise Researchers

ASIRPA RT is a formative evaluation tool with the aim of accompanying researchers to envision desired futures and to navigate research in that direction (Matt et al., 2023). The

approach was implemented in the PPR-CPA, the French mission-oriented research programme with the aim to eradicate pesticides by 2030-2040. In this way, the funded researchers were guided in considering their contributions to pesticides eradication and the societal impacts that such eradication might bring. We compared researchers' visions of eradicating pesticides before their participation in ASIRPA RT (T0) (analysis of proposals and interviews) and one year after (T1) (analysis of data gathered during workshops and activities).
ASIRPA's central tool is the 'impact pathway' (IP), which describes the non-linear process of how scientific knowledge is expected to be translated into five dimensions of societal impacts. The approach mobilises the Sociology of Translation (Callon, 1986), to highlight the chains of translation that occur during the process of generating societal impacts. Through four moments, Callon's (1986) translation model describes how such chains of translation mobilise a heterogenous network, as researchers negotiate and enrol other actors in contributing towards a predefined problem.
The particularity of a formative approach in real-time, like ASIRPA RT, is that it involves anticipating contributions to an envisioned future while the R&I activities are still ongoing. If we aim to understand how ASIRPA RT guides the responsibilisation of researchers in their contribution to a societal mission, we thus need to understand the anticipation of translation processes. Therefore, our conceptual starting point in this regard is that achieving a mission requires change in society, and this societal change is the result of the mobilisation of actors over time in heterogenous networks. This means that in the case of the PPR-CPA, the researchers have to (re)think their own roles with other actors' roles in heterogenous networks. It is this network that will put in practice the studied alternative solutions to pesticides, and thus enable the researchers' envisioned contribution to a future 0-pesticides society. These networks are constructed around visions of responsibilities different actors hold in their different roles to perform the eradication of pesticides in society. For understanding the progressive construction of these envisioned networks, we are building upon the 4-dimensional framework of Responsible Innovation (RI) by Stilgoe et al. (2013): anticipation, reflexivity, inclusion and responsiveness.

A Change in Envisioning Science-Society Associations:
From 'Science for' to 'Science with' Society ASIRPA RT accompanied the PPR-CPA researchers to rethink the associations between their research and the type of future society that is needed for their results to work, and to take on responsibility in this regard. After providing an analysing of the T0-phase (before ASIRPA RT), we illustrate how the

researchers assumed responsibility as they started to anticipate the process of translation with ASIRPA RT during the T1-phase. Our analysis highlights that at the T0-phase the researchers demonstrated to have renewed and more ambitious scientific questions, than in classical R&I projects. The researchers also added complexity to their projects on various levels such as: multi-disciplinary and multi-actor processes, multiple scales from the laboratory to the field, and by addressing multiple pests or cropping systems. However, the research proposals emphasise their focus on excellent science, which reflects visions of their contribution to the mission through a scientific way of approaching how a society without pesticides should be constituted and act. It shows visions of a straightforward interest, adoption and use of alternative solutions by users in society, for which it is the task of research to provide scientific proof of their efficiency. Despite the evident ambition of the research projects to contribute to a societal mission, the researchers' capacity to envision societal change enabling the eradication of pesticides with the alternative solutions is rather weak if they are not guided in that direction.

At the T1-phase, the PPR-CPA researchers began to envision their anticipation, reflexivity, inclusion and responsiveness towards chains of translation that are necessary in bridging the research on alternative solutions and a society that will use these solutions to eradicate pesticides. This process is based on three visions of change that the researchers emphasised for a future society without pesticides:

1. Visions of transformations – The researchers shared their visions of the constitution of a future society that has eradicated pesticides thanks to the alternative solution they will have developed. Researchers began to articulate the required transformative change in society, which are expected to take place in five interdependend poles: Technology, User context, Market, Policy & Regulation, Science & Education.
2. Visions of actors – The researchers' descriptions of societal transformations embed visions of the actors who should be responsible for that change. It reveals that they expect to be able to eradicate pesticides once other actors develop further and embed the researchers' alternative solutions into their practices.
3. Visions of the state of the future network – The researchers started exploring strategies for ensuring that other actors will take on their new responsibilities, which are envisioned by the researchers in imagined networks that are to perform chains of translation. These associations become real as the researchers begin to anticipate how they can raise interest and enrol actors.

By exploring these three visions of change, it becomes apparent how inextricably linked responsibility and translation are. We show that chains of translation deal with both the visions of the responsibilities of researchers and other actors and how these might reconfigure a future society. In developing these visions of change, it required the researchers to anticipate the directionality of their research to a possible future society, to reflect on responsibilities of other actors, to include those actors in R&I processes, and to be responsive in (re-)directing R&I pathways towards societal goals. Consequently, our findings suggest that the researchers assumed responsibility when they associated the visions of responsibilities in the network, so that the alternative solutions can speak for the network that will put them into practice in the future. Put differently, responsibilisation is about associating actors in networks based on the responsibilities they should assume in order to perform (change towards) the future society that has achieved the mission.

Hence, as the PPR-CPA researchers were accompanied by a formative evaluation approach, they started anticipating the construction of their envisioned networks that are needed to translate their research on alternative solutions to a future society that has eradicated pesticides. It should be noted, though, that these visions reflect an analysis of researchers only in an early phase of the projects. Over time, visions need to further advance, and this process should become more participatory and inclusive with other actors.

As the researchers started to anticipate visions of change when they were accompanied by ASIRPA RT, they no longer envisioned their contributions to the eradication of pesticides through a linear approach to innovation (transfer of research outputs). Instead, the researchers started anticipating the roles and responsibilities of actors within an evolving network (translation of research results). Hence, through their associations in the networks, the researchers are realising that it is only by envisioning the sharing of responsibilities with other actors that their alternative solutions will be able to eradicate pesticides. This means that the science-society associations are represented by these networks.

The four-dimensional framework of RI by Stilgoe et al. (2013) helps to understand the construction of the envisioned networks, as normative visions of responsibility. By differentiating among responsibilities that act in the networks, we showed that these values are captured in visions of change of a 0-pesticides society via new associations in the future networks that should put in practice the alternative solutions. These associations are made through the chains of translation. Our results demonstrate therefore that these four-dimensions (embedding visions of change) should become part of the translation process (i.e., they should be translated) in order to be enacted. Finally, we reflect upon how our results might provide a promising way to institutionalise responsible research for mission-oriented policy contexts.

REFERENCES:

- Arnaldi, S., & Gorgoni, G. (2016). Turning the tide or surfing the wave? Responsible Research and Innovation, fundamental rights and neoliberal virtues. Life Sciences, Society and Policy, 12(1), 1–19.
- Callon, M. (1986). Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay. The Sociological Review, 32(1), 196–233.
- Dorbeck-Jung, B., & Shelley-Egan, C. (2013). Meta-Regulation and Nanotechnologies: The Challenge of Responsibilisation Within the European Commission's Code of Conduct for Responsible Nanosciences and Nanotechnologies Research. NanoEthics, 7(1), 55–68.
- Douglas, H. E. (2003). The Moral Responsibilities of Scientists (Tensions between Autonomy and Responsibility). American Philosophical Quarterly, 40(1), 59–68.
- Giddens, A. (1984). The Constitution of Society Outline of the Theory of Structuration. University of California Press.
- Jasanoff, S. (2004). States of Knowledge: The Co-Production of Science and the Social Order. Routledge.
- Joly, P.-B. (2018). Innovation and the Problem of Value. NOTE DE RECHERCHE IFRIS N6.
- Kuhlmann, S., & Rip, A. (2018). Next-Generation Innovation Policy and Grand Challenges. Science and Public Policy, 45(4), 448–454.
- Latour, B., & Woolgar, S. (1979). Laboratory life : the construction of scientific facts. SAGE Publications.
- Matt, M., Robinson, D. K. R., Joly, P.-B., Van Dis, R., & Colinet, L. (2023). ASIRPAReal-Time in the making of how to empower researchers to steer research towards desired societal goals. Research Evaluation, 32(2), 412–425.
- Robinson, D. K. R., & Mazzucato, M. (2019). The evolution of mission-oriented policies: Exploring changing market creating policies in the US and European space sector. Research Policy, 48(4), 936–948.
- Schot, J., & Steinmueller, W. E. (2016). Framing Innovation Policy for Transformative Change: Innovation Policy 3.0.
- Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. Research Policy, 42(9), 1568–1580.

- Strydom, P. (1999). The challenge of responsibility for sociology. Current Sociology, 47(3), 65–82.
- Temri, L. (2008). Responsible innovation. Vincent, N. (2010). A Structured Taxonomy of Responsibility. In N.
- Vincent, I. van de Poel, & J. van den Hoven (Eds.), Moral Responsibility (Library of, pp. 15–35). Springer Dordrecht.

Georg Mildenberger, Achim Oberg, Maria Rabadjieva, Elisa Sauerbier, Judith Terstriep, Dominika Wruk, Filip Zieliński
Centre for Social Investment, Heidelberg University, Germany
Center for SME Research and Entrepreneurship & University of Hamburg , Germany
Westphalian university of applied sciences, Institute for work and technology (IAT), Germany
Center for SME Research and Entrepreneurship, Mannheim University, Germany
Westphalian university of applied sciences, Institute for work and technology (IAT), Germany
Center for SME Research and Entrepreneurship, Mannheim University , Germany
Centre for Social Investment, Heidelberg University , Germany

EVALUATING IMPACTS OF SOCIAL INNOVATIONS WITH FIELD-SPECIFIC MODELS

ABSTRACT

The research and practice field of Impact Evaluation and Social Impact Measurement remains fragmented into numerous tailor-mades, one-off solutions on the one hand and very abstract models with general indicators on the other hand. The same holds true for the measurement of the impacts of innovations and - more specifically - Social Innovations (SI). Here, the ISI Project4 tries to fill the gap in-between the macro-level of social change and the micro-level of SI actors. To reach this goal, we developed a middle-range impact model and indicators that can be applied within specific fields (communal living, digital education, blockchain, sharing economy) of SI.

As part of the ISI project, we have developed a conceptual framework that allows us to integrate SI into the more general concept of innovation in innovation research, on the one hand, and impact evaluation research, on the other hand. Specifically, we suggest positioning the various existing understandings of SI within three overlapping frames, or dimensions, of innovation: 1.) The continuum of tangible vs intangible (=”social”) types of innovation objects; 2.) the continuum of bottom-up (=”social”), participative innovation vs top-down, centralised innovation; 3.) the various value orientations that underlie the intentions to innovate, with the general distinction of for-profit and non-profit (=”social”) orientations. If innovation is defined as a new and thus (at least potentially) better solution to existing problems, it clearly includes a causal relationship of causes (the “solution” or “innovation object”) and effects (the problem-solving). It is the third frame, i.e. the social (including: environmental) intention, that allows us to define SI in such a way that it can be linked with impact measurement, which aims to provide evidence to validate assumptions of causal chains for organizations, programs or other interventions. In short, the intention defining a given SI can be framed as a Theory of Change, thus making use of existing approaches in evaluation research and practice.

the intention defining a given SI can be framed as a Theory of Change, thus making use of existing approaches in evaluation research and practice.

Three major approaches in impact measurement can be differentiated:

1. Monetization-based valuation (CBA, SROI, Impact Accounting), a dynamically developing field of impact research, offers standards to translate impacts into a common unit (money) to make them comparable and easier to communicate. However, it often relies on existing approaches, such as the Theory of Change (ToC). It leaves valuation to market mechanisms where possible, reverting to rate fixing for ethical reasons in some cases (e.g. value of human life vs death).
2. Theory-based approaches in impact evaluation include, for example, Program Theory, Rationalist Evaluation, Contribution Analysis, Process Tracing, Most Significant Change, QuIP and QCA (Weiss, 1997; Rogers, 2000). They focus on identifying causal chains and understanding how, for whom and in what contexts change is caused. Most importantly, causal inference, here, means checking (verifying/falsifying) the assumptions of the ToC against evidence in order to refine these assumptions. Generally, the differentiation between outcomes and impacts is made from the point of view of the ToC of the intervention/program/organization that can be differentiated in actors and other objects; causal chains; time and space; intentionality; visibility and evaluation.
3. Variance-based approaches, i.e. RCTs, Propensity Score Matching, Instrumental Variables, Difference-in-Difference, Regression Discontinuity (see Gertler et al., 2016 for an overview), and current updates made possible by Machine Learning (Brand et al., 2014), on the other hand, start with a different understanding of causality and thus definitions of outcome and impact. Here, causal effects are understood as not directly observable but correspond to (average) differences between variables of change for a sample of entities, compared to a control group that serves as a representation of the counterfactual. From this perspective, “impact” can be defined as those changes (=outcomes) of an intervention, program, organization, etc., that would not have happened without it.

We recommend a pragmatic, theory-based, participatory approach for SI impact measurement that allows for mixed-methods (quant and qual) designs; but to search for ways to enhance it in a way that answers the important questions of attribution and impact intensity based on a counterfactualbased definition of outcome and impact. More specifically, our proposed “model for social innovation impact measurement” (MSIIM) is based on the ToC method, including the well-established InputOutput-Outcome-Impact causal chain logic (originating in the logical framework approach), risks and contextual factors for each link, and indicators to check the assumptions of the ToC against evidence (quantitative and standardized as much as possible). To gain the needed understanding, we combined desk research with semi-structured interviews with social innovation practitioners in each field. The data collection allowed us to identify the essential characteristics for the development of impact models, e.g. Borders, Social Innovation, available statistics and organisational forms of SI. We will exemplify the development process of such MSIIMs by presenting two innovation fields: digital education and blockchain.

Social innovation is not a new phenomenon in the area of education (Maldonado-Mariscal & Alijew, 2023). In the past decades, however, digitalisation has opened the door for new educational practices and has become a sub-area for social innovation where a variety of technologies have been utilised for educational purposes (Zawacki-Richter & Jung, 2023). In the ISI project, we define social innovation within digital education as technologically driven solutions that enable the general public or sub-groups to participate in education in ways they could not have without these technologies. Specifically, we focus on three categories: Massive Open Online Courses (MOOC), Open Education Resources (OER) and Applications (Apps). The organisational forms, participants and beneficiaries in the three categories vary. For example, MOOCs are typically organised by educators or teachers at educational institutions. However, to participate in MOOCs, students need to register via a specialised platform that manages the MOOC process, usually from content development through participation until certification. This constellation results in the identification of various roles that need to be considered for the impact measurement. In contrast, Apps are usually organised as companies and present a centralised business model with generally two roles - the app as a provider and the customers as users. To assess the impact of the mentioned social innovations on education and society, we identified specific indicators that can be measured directly by questioning the SI organisers and participants or combined with existing metrics. Such indicators include, among others, organisational capacities (i.e. employees, working hours and other necessary resources for creating and participating in the solutions) and quantification of engagement (i.e. number of visits, interactions, content management etc.), allowing to build casual chains from input to impact. The second example in which we attempt to pave a way towards impact measurement is blockchain technology. In his 2008 whitepaper, Satoshi Nakamoto described blockchain as a peer-to-peer electronic cash system that would allow transactions between two parties without the need for an intermediary (Nakamoto, 2008). Ever since this first publication, blockchain technology has come a long way, not only changing the financial services sector but also allowing new ways to store data, safeguard assets, secure identities and collaborate (Sandner et al., 2019; Tapscott & Tapscott, 2016). The principle on which all these new forms of application are based at their core is decentralisation in conjunction with smart contracts, which guarantees security without the need for a third trusted party, rendering institutions redundant in specific contexts. Within the ISI project, we define the borders of the field by focusing on adopting a distributed ledger technology for disintermediation. However, recurring patterns in the different types of organisations enabled us to identify certain roles. In terms of possibilities for impact measurement, there are public databases available from which information on different blockchain-based transactions and organisations can be retrieved (e.g., DeepDAO, DefiLlama, CoinMarketCap). In terms of applications with a stronger social or environmental focus, impact tokens that combine fundraising for a good cause and investor value creation are directly linked to specific outcomes, quantified by metrics such as the number of vaccinations administered or tons of carbon dioxide mitigated. Yet, the assumed impact that blockchain technology unfolds goes far beyond available sources, so we created an elaborate indicator selection comprising roles and direct outcomes for participants, employees and users.

The advantages of a standardized approach are manifold. First, results could be compared and aggregated for the benefit of innovative actors, funders and research. Second, it would be easier to train professionals to conduct impact evaluations. Third, organization leaders, funders and the general public could understand the reports and their results with less effort. On the downside, such a standardized methodology will necessarily ignore important differences in types of activity and societal contexts as well as idiosyncrasies of problems and target groups with potentially adverse societal effects. To some extent, this can be mitigated by including a strong participative component in the model development design and by a thorough understanding of the SI-fields at hand.

Furthermore, we argue that this work of mapping SI-fields, developing generic models, collecting, aggregating and analyzing data cannot reasonably be done by SI actors alone but only in collaboration with the social sciences due to the resources, knowledge and objectivity that are needed. To enable measurement practitioners and interested actors to use the suggested model, we want to discuss the challenges of field definition, data collection, and analysis of further social innovations.

Our work has shown that even though social innovation fields may be very different in the object of innovation, the process of defining the field and respective indicators are similar. Furthermore, following the same IOOI logic, it is possible to build field-specific causal

chains. For example, social inclusion is considered a desirable impact in both presented SI examples and similar metrics can be used.
Only with reliable impact evaluation is it possible to assess whether this intention of inclusion, for example, is reflected by evidence, and consider positive and negative, intended and unintended, direct and indirect, short-term and long-term effects. Field-specific models can help to make this ambitious task somewhat more practicable.

REFERENCES:

- Brand, Jennie E.; Zhou, Xiang; Xie, Yu (2023): Recent Developments in Causal Inference and Machine Learning. In: Annu. Rev. Sociol. 49 (1), S. 81–110. DOI: 10.1146/annurev-soc-030420-015345.
- Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. (2016). Impact evaluation in practice. World Bank Publications.
- Maldonado-Mariscal, K. and Alijew, I. (2023): Social innovation and educational innovation: a qualitative review of innovation's evolution, Innovation: The European Journal of Social Science Research, DOI: 10.1080/13511610.2023.2173152.
- Nakamoto, S. (2008). Bitcoin whitepaper.
- Rogers, P. J. (2000). Causal models in program theory evaluation. New directions for evaluation, 2000(87), 47-55.
- Sandner, P., Welpe, I., & Tumasjan, A. (2019). Der Blockchain-Faktor: Wie die Blockchain unsere Gesellschaft ver‰ndern wird. BoD – Books on Demand. Tapscott, D., &
- Tapscott, A. (2016, May 10). The Impact of the Blockchain Goes Beyond Financial Services. Harvard Business Review. https://hbr.org/2016/05/the-impact-of-the-blockchain-goes-beyondfinancial-services
- Weiss, C. H. (1997). Theory-based evaluation: past, present, and future. New directions for evaluation, 76, 41-55.
- Zawacki-Richter, O. and Jung, I. (2023): Shaping the Field of Open, Distance,and Digital Education. In: Zawacki-Richter, O., Jung, I. (eds) Handbook of Open, Distance and Digital Education. Springer, Singapore. https://doi.org/10.1007/978-981-19-2080-6_94.

Constanze Fetting
Centre for Social Innovation (ZSI), Austria

MONITORING TRANSFORMATIVE ADAPTATION FOR SOCIO-TECHNICAL ECOLOGICAL SYSTEM CHANGE

ABSTRACT

As the impacts of climate change become impossible to ignore, climate change adaptation is rising in importance on the agenda of governments including research funders. As one of the five missions of the European Union, 'Adaptation to Climate Change', the topic is also gaining traction on the European level, and several research and innovation projects, including various that are funded through the Horizon Europe programme, have recently taken off. R&I for climate change adaptation comes with a unique set of challenges. The greatest among them is the risk of maladaptation, which would leave the involved communities more vulnerable than before an intervention. Effective monitoring and evaluation can be a useful tool in preventing maladaptation, ensuring a just transition, and embedding the intervention in the socio-technical ecological system.

This contribution presents a Monitoring Framework for transformative climate change adaptation in a socio-technical ecological system (STES). The Horizon Europe project MountResilience conducts six regional climate change adaptation R&I interventions across Europe with the aim of increasing regions' climate resilience and increase adaptive capacity in regional actors.

The Theory of Change combines the STES approach with the concept of transformative adaptation. Consequently, the monitoring framework covers all dimensions of the regional STES (the social, technological/practice, and ecological dimensions). Leaning on the concept of transformative adaptation and innovation, it encompasses and addresses the underlying drivers of vulnerability, encouraging systemic change and long-term resilience, while at the same time preventing maladaptation. The monitoring framework tracks and guides all these interconnected aspects to ensure consistent mission orientation and progress towards the desired transformative impacts. The contribution thus aligns with the goals of Track #1: Evaluation of transformation policies and dynamics in socio-technical systems as it expands the STS concept with a clear focus on sustainability by including the ecological component. At the same time, it emphasises transformative outcomes and societal wellbeing.

As the framework aims at deep systemic changes, it naturally links the R&I intervention to other policy areas, such as ecosystem management, health, agriculture, and energy. For systemic change to materialise, the intervention naturally must impact a wide array of policy areas. While we are not directly monitoring transformative R&I policies, the interventions are designed to have impacts beyond their direct target and are likely to trigger policy change from below and across silos. STES frameworks display the interdependencies, feedback loops, and dynamics between social, technical, and

ecological components within a given context, in our case a regional system. Through this holistic approach,

STES frameworks provide valuable insights into how changes in one system can impact other components, thus offering a comprehensive understanding of complex socio-technical-ecological phenomena. Transformative climate change adaptation, addressing the systemic drivers of climate vulnerability, requires a wholeof-system approach to include ecological and societal drivers as well as the intervening force of technologies or social practices that mediate the interaction between ecological and human systems. This contribution will thus present an approach to monitor not only the three STES dimensions, but also the interactions between them. It addresses the complex system criterion which supports change within times of turbulence, uncertainty, and uncontrollability. The STES monitoring approach provides a method to assess the interactions between the different components of a complex system and, if necessary, adapt the R&I process to new regional circumstances or multi-level developments in each of the dimensions. While the monitoring is mostly qualitative, the goal of the R&I intervention is to increase both societal and ecological resilience. Therefore, it will be important to include biophysical elements into the monitoring and evaluation framework. This will be part of monitoring the ecological dimension, together with questions of ecosystem management and ecosystem services valorisation.

Transformative adaptation aims at fundamentally changing the constituents of a system that render it vulnerable, such as its institutions, structures, and practices. It should not only directly address the impacts of natural hazards (which is also called coping or incremental adaptation) but deal with the underlying drivers of the system's vulnerabilities and lack of resilience. It should prevent maladaptation, which may lead to short-term resilience but would leave the system more at risk in the long term. One objective of transformative adaptation is to create and strengthen transformative capacities in all stakeholders in the region (i.e. going beyond government actors).

In MountResilience, we use the concept of Three Spheres to ensure that the results of the R&I activities are embedded in the regional STES. This tool to conceptualise the mechanisms of societal transformation is based on a dynamic systems view.6 The practical sphere includes behaviours and technical responses to climate change. It is thus where (social) innovation is located. In the political sphere we monitor changes in systems and structures, including norms, rules and institutions. In climate change adaptation, this also includes ecosystem management, as that is often a political or administrative task. Adaptive capacities, which describe the abilities of actors to react to, mitigate and prevent the impacts of climate change in the region, are usually embedded in institutions. Finally, the personal sphere includes subjective beliefs, values, and worldviews. This will be monitored partially through surveying the population in each region and assessing eco-system services. It is necessary to include actors from all three spheres to embed the results of the R&I activities in the regional system and to make transformation successful. Consequently, the monitoring framework will need to observe and evaluate interactions and impacts of the innovation with actors in all three spheres. This includes guiding the innovation partners in reaching out to and embedding their activities, for example through targeted stakeholder engagement. Through reaching out to actors in the Three Spheres and the three dimensions of the STES, the monitoring framework ensures a holistic approach and can observe and guide the development of adaptive capacities in actors from all dimensions of the regional system.

As interventions within the ecological and climate realm are characterized by uncertainty, we build on a developmental evaluation methodology that is built on flexibility and can be adapted to new and emerging circumstances and developments. This is especially important as actions in a complex system can lead to unintended consequences. We thus present a systemic approach that goes far beyond the STS framework by not only adding the ecological dimension, but also the potential transformative effects of an R&I intervention. R&I activities in the realm of climate change, be it mitigation or adaptation, require a complex systems approach, encompassing many different elements and the interactions between them. This contribution will not only present those different elements, but also an approach to guide the leaders of the intervention to stay on track towards the desired transformative vision through a close-knit and expansive monitoring approach.

Erika Hajdu, Dr Giovanna Lima, Dr Stefan de Jong
Erasmus University Rotterdam, Netherlands

DESIGNING A THEORY OF CHANGE IN LESS THAN AN HOUR

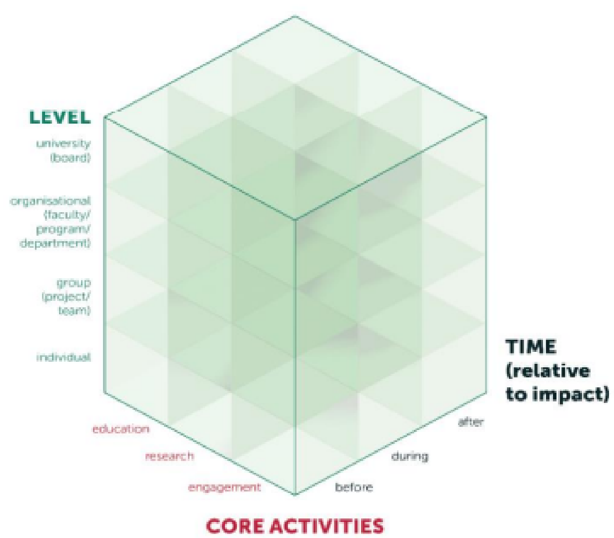
ABSTRACT

The past decade has intensified the focus on societal impact within universities, encouraging research(ers) to address grand societal challenges in collaboration with partners from other disciplines (interdisciplinarity) and beyond academia (transdisciplinarity). Facilitating the integration of these multiple epistemic communities, with diverse perspectives and lived experiences, around a shared common vision and with concrete objectives and evaluation criteria of what success looks like has so far remained a challenge (LERU, 2023).

Academic organizations are transforming their way of working to be more responsive to societal challenges; this shift requires new structures towards societal impact as a process that enables taking shared public responsibility, involving learning and reflection, instead of solely focusing on results and research excellence based on scientific and commercial impact (see D'Este et al., 2018). Many frameworks and methods have been proposed that provide the theoretical background to plan, enact, evaluate, and demonstrate one's positive contribution to society (e.g., SIAMPI (Spaapen & van Drooge, 2011)) including contributions from the design field (see Design for Social Innovation, Transition Design (Irwin et al., 2020).

However, the use of theoretical methods across disciplines falls behind their potential. Application proves to be challenging in the academic context riddled with rigid structures and evaluation cycles, and obstacles regarding capacity, financial- and time constraints, and competing priorities (Hughes et al., 2016). To facilitate impact activities to flourish across our institutions, we need to embed these methods and practices into the academic context and thus create cultural change on various levels (Figure 1), requiring distinct types of support (Perkmann et al.,2019).

FIGURE 1- REPRESENTATION OF THE DIFFERENT DIMENSIONS OF IMPACT WITHIN UNIVERSITIES



Within the context of research projects, however, we see Theories of Change (ToC) (e.g Belcher & Claus, 2020) emerging as a main method, now widely used in evaluations (Mayne, 2017) and in research proposals, like the Impact Pathway in Horizon Europe application forms (EU, 2024). While this method has been increasingly preferred, and workshops on developing ToCs are extremely popular, our experience has shown that the development of a ToC was still perceived as too much time and effort and felt as an additional chore to the research teams.

Given this scenario, the team Evaluating Societal Impact (ESI) at Erasmus University Rotterdam (EUR) has taken a user-centred approach to translating the method into a hands-on tool for the context of inter- and transdisciplinary projects: the Journey of Progress – Theory of Change card game. It addresses the unique challenges of such types of transformative collaborations, such as stimulating co-creation from within while avoiding fatigue of partners and negotiating one's span of control. The tool empowers and enables projects, initiatives, and various organisational units to autonomously build their ToC.

Using the Journey of Progress card set, members of a group take a first step towards formalising their understanding on how and why change is supposed to happen in their given context. By backcasting and thinking collaboratively about a joint goal and ways to get there, participants build pathways from their desirable future to the specific actions needed to realise the necessary steps towards making change. The card set can be used in various contexts to understand envisioned change processes.

In addition to the physical card set, two short introductory videos about ToC and about the game itself are available to facilitate its use. In situations in which all partners cannot be physically present, a digital version is available within an online collaborative environment (Miro). The tool also can be supported with interviews that allow for the perspectives of those who cannot be in the (virtual) room: they still get represented in the design of the shared ToC.

The standard game tackles the basics of building a ToC in only 40 minutes, including a vision statement, outcomes, outputs, inputs, and activities. For those wanting to refine the results of the standard session, further extensions are available such as cards for defining stakeholders, or bringing specific assumptions to the surface. Another extension of the game helps users to develop their indicators for monitoring and assessing their planned changes. These extensions are particularly important for the purpose of the REvaluation Conference on assessing and measuring transformation policies and measuring social innovation.

FIGURE 2 - JOURNEY OF PROGRESS (PROTOTYPE VERSION PICTURED), BY THE EVALUATING SOCIETAL IMPACT TEAM AT ERASMUS UNIVERSITY ROTTERDAM

The tool has been extensively trialled and tested across multiple types of projects, with a variety of stakeholders (including various academic and administrative organisational units within the university, municipalities such as Amsterdam and Rotterdam, and governments and government agencies, such as the Dutch Ministry of Science and the Dutch Research Council, leading to positive outcomes in terms of quickly developing a shared vision in complex environments and overcoming power differences between participants. International interest is gaining momentum. For instance, a gaming session is requested by the University of Cambridge. The tool has been highly sought after, with colleagues from higher education institutions and public research organizations across the Netherlands wishing to purchase the tool which indicates its perceived usefulness by users.

We see a strong fit with Strand 1. The Journey of Progress, as it is true with any ToC tools, can be used to develop relevant indicators to measuring social innovation within transformative R&I policies, and can be used in multiple areas. When compared to other tools, the main benefits are: 1) it is context-independent; 2) autonomously deployed with no previous knowledge; 3) time saving; 4) designed and trialled by and with academic and societal partners.

We understand many in the REvaluation conference may benefit from using the tool. We would be delighted to not only present the tool but also add information about its rationale, limitations and about how it has been developed and our lessons learned in tool development for enabling impact. This could be done in a stand-alone session of about 1 hour, in which participants get to trial the tool themselves, or via a shorter 30 min presentation followed by Q&A. We welcome the suggestions of this peer-review process of academic and evaluation experts.

REFERENCES:

- Belcher, B., & Claus, R. (2020). Theory of change. Td-net toolbox profile (5). Swiss Academies of Arts and Sciences: td-net toolbox for co-producing knowledge. www.transdisciplinarity.ch/toolbox. doi.org/10.5281/zenodo.3717451
- D'Este, P., Ramos-Vielba, I., Woolley, R., & Amara, N. (2018). How do researchers generate scientific and societal impacts? Toward an analytical and operational framework. Science and Public Policy, 45(6), 752–763. https://doi.org/10.1093/SCIPOL/SCY023
- European Comission (2024). Horizon Europe Programme: Standard Application Form (HE RIA, IA). https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/temp-form/af/af_heria-ia_en.pdf
- Evaluating Societal Impact. (2024). Journey of Progress – Theory of Change card game. Erasmus University Rotterdam (EUR). https://doi.org/10.25397/EUR.25311478.V1 Evaluating Societal Impact. (2024). Introduction to Theory of Change. YouTube video. https://youtu.be/wa5-p9hcGc0?si=OFcq2MUhUyNj-TQq Evaluating Societal Impact. (2024). How to play Journey of Progress. YouTube video. https://youtu.be/YQFCySkvp-0?si=egPOb-w6KajC7WIA
- Hughes, A., Lawson, C., Salter, A., Kitson, M. with Bullock, A. and Hughes, R.B. (2016). The Changing State of Knowledge Exchange: UK Academic Interactions with External Organisations 2005 -2015. NCUB, London.
- Irwin, T., Tonkinwise, C., & Kossoff, G. (2020). Transition design: An educational framework for advancing the study and design of sustainable transitions. Cuadernos del Centro de Estudios en Diseño y Comunicación. Ensayos, (105), 31-72.
- LERU (2023). Implementing interdisciplinarity in research-intensive universities: good practices and challenges. Advice Paper no. 30. March 2023. https://www.leru.org/publications/implementinginterdisciplinarity-in-research-intensive-universities-good-practices-and-challenges
- Mayne, J. (2017). Theory of Change Analysis: Building Robust Theories of Change. In Canadian Journal of Program Evaluation (Vol. 32, Issue 2, pp. 155–173). University of Toronto Press Inc. (UTPress). https://doi.org/10.3138/cjpe.31122
- Perkmann, M., Salandra, R., Tartari, V., McKelvey, M., & Hughes, A. (2021). Academic engagement: A review of the literature 2011-2019. Research Policy, 50(1), 104114. https://doi.org/10.1016/J.RESPOL.2020.104114
- Spaapen, J., & van Drooge, L. (2011). Introducing 'productive interactions' in social impact assessment. Research Evaluation, 20(3), 211–218. https://doi.org/10.3152/095820211X12941371876742

Diogo Machado
Technopolis-Group, France

THE MEANING OF SCIENTIFIC NOVELTY: LEARNING FROM EXPERT REVIEWS

ABSTRACT

Lack of risk-taking in scientific research is weighing on our capacity to generate breakthrough discoveries. We are in desperate need of highly novel and transformative discoveries that can help us tackle grand societal challenges like climate change, ageing, and global pandemics and boost productivity. Despite this need, existing tools to effectively inform, monitor and evaluate the capacity of policies to foster transformative research are notably limited. The lack of robust indicator frameworks for transformative research gives traditional bibliometric indicators an excessive prominence which makes matters worse. Traditional bibliometric indicators based on impact-factor metrics have been shown to be biased against highly novel, transformative research [see e.g. Wang et. Al (2017), OECD (2021) and Machado (2021)].

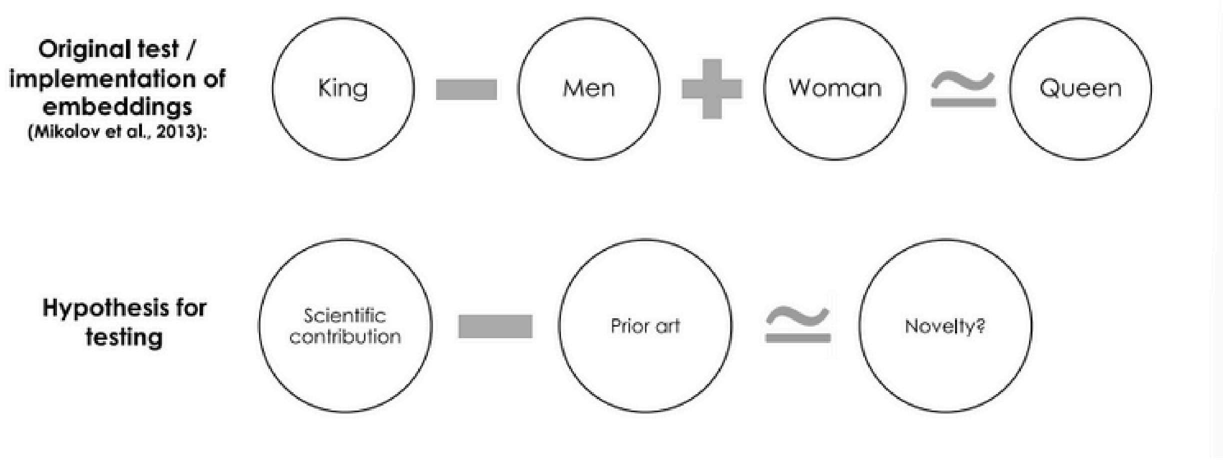
This paper contributes to the scholarly literature dedicated to refining our understanding of the intricate nature of scientific novelty. Franzoni and Stephan (2023) acknowledge how challenging this research area is, encouraging others to take up this important topic. In this paper, I do so by investigating new avenues capable of measuring and predicting how novel new scientific ideas are. This endeavour is not just academic; it holds potential practical significance for science policy. Tools that help identify highly novel ideas have substantial potential to enhance our understanding of the factors that drive novelty and transformational research. Furthermore, this research paves the way for developing innovative tools designed to support peer review processes, streamline the detection of potentially highly novel ideas and help science funders monitor the novelty levels of their portfolio of investments. Such advancement can benefit the broader scientific landscape at several stages, such as project selection, publication processes, and science policy monitoring and evaluation.

The paper makes multiple novel contributions to the literature. I developed a new framework to measure novelty which learns from expert reviews of top scientists, instead of simplistic assumptions or crude proxies. The framework is flexible, adaptable to different fields and more comprehensive to better capture the complex dimensions of novelty and transformational research. Moreover, a limitation of most bibliometric indicators is that they only apply to published articles, not capturing the potential of new scientific ideas at the proposal level and requiring ex-post information from after articles' publication date. In contrast, the proposed framework measures novelty potential at the idea conception stage (proposal level), but can also measure the novelty content of published articles. Finally, this paper is the first to apply a combination of machine learning methods (random forests) and genetic algorithms to full-text data from funded and unfunded scientific proposals in all fields of science.

The major innovation of this paper is the development of a "novelty vector". Consequently, the central hypothesis for testing is whether the vector manages to learn from expert reviews and predict novelty using a unique dataset of successful and unsuccessful scientific proposals and their expert review scores. This vector stems from a new type of data exploration using textual embeddings. Embeddings are a form of textual representation originating from machine learning and natural language processing disciplines, which captures semantic meaning based on the context in which textual items appear in large corpora. Embeddings translate words, phrases or documents into vectors of real numbers. These vectors fit a multi-dimensional space such that documents with similar meanings or contexts point at similar directions, while those with distinct meanings are further apart.

The "novelty vector" results from exploring untested characteristics of embeddings inspired by one of the first discoveries of embeddings' potential to understand semantic meaning. In embeddings' original proposal, Mikolov et al. (2013)6 show that the subtraction between the vector for the textual terms "King" and "Men" followed by the sum of a vector for "Women" results in a vector with high similarity to the vector for "Queen". To a large extent, this finding formed the conceptual backbone behind current developments in Generative AI and large language models. Inspired by this finding, I computed the "novelty vector" by subtracting the vector for a focal proposal and the vector representing its prior art. The intuition is that when we remove the semantic meaning of the prior art to a certain scientific advancement, the result represents what is novel. Consequently, the indicator framework exploring the "novel vector" can open a promising area of applications for novelty assessment and transformational research policy.

FIGURE 1. CONCEPTUAL INSPIRATION OF THE NOVELTY VECTOR



The paper thus tests the new indicator framework by exploring the text of scientific proposals and the titles and abstracts of their respective prior art. Different dimensions of the "novelty vector" enter a genetic machine learning algorithm – a search heuristic inspired by the process of natural selection – using these dimensions to predict the component of the expert review scores that specifically assesses proposals' novelty. When using the results from the model to predict a testing dataset of completely unseen data to the model, these dimensions manage to predict reviewers' top novelty scores with a precision of 96.4%. Moreover, further analysis shows that this framework is relevant to novelty even when holding constant the quality of the proposal and the strength of applicants' CVs (measured through further expert review scores) and only comparing proposals in the same year and scientific field. Multiple robustness checks show that these findings are not likely to result from an artificially pervasive capacity of the genetic algorithm to find any relevant prediction pathways. Namely, following precisely the same methodology to predict 1) the final funding decision and 2) other expert review scores (feasibility of the approach) results in prediction accuracies below 30% and a lack of joint statistical significance under the same econometric framework. These tests demonstrate that the "novelty vector" is specifically suitable for measuring originality/novelty, not other knowledge dimensions. To test external validity, I applied the framework to a dataset of published papers (instead of proposals). The dimensions of the novelty vector were statistically significant when identifying Nobel prize-winning papers even after accounting for forward citations (as a proxy for ex-post quality), journal/issue (as a proxy for ex-ante quality filter), and topic similarity.

Matthijs Kallenbach, Matthijs J. Janssen
Copernicus Institute of Sustainable Development, Utrecht University, Utrecht, The Netherlands

EVALUATING THE ROLE OF POLICY IN TRANSITION MECHANISMS: THE SUCCESS CASE OF THE ELECTRIC BUS IN THE NETHERLANDS

ABSTRACT

Addressing urgent societal challenges provoked by climate change requires drastic reconfiguration of socio-technical systems, which create winners and losers among affected populations and regions (Geels & Schot, 2007; Upham et al., 2022). Research in policy sciences and transition literatures discusses how public organisations can act upon these issues, notably by adopting governance arrangements that allow them to create and adapt coherent policy mixes (Kivimaa & Kern, 2016; Kern et al. 2019; Borr·s & Edler, 2020; Braams et al., 2021). Despite the increasing coming together of these scholarly fields, a long-standing criticism of policy research is that it fails to properly take into account socio-technical system dynamics (Alkemade & Dekoninck 2021; Haddad et al. 2022). Hence, we still have a limited understanding of key policy-driven transition mechanisms, in which policies interactions with a variety of other systemic processes. This research asks the questions: what are key policy-driven transition mechanisms and how can we build an evaluation framework to identify them?

Identification of policy-driven transition mechanisms requires going beyond the status quo in transition policy evaluation, where the focus is on assessing the performance of specific policy programs. Current most advanced evaluation methods can be found in the literature on “Transformative Innovation policy” (Borrás & Edler, 2020) which considers policy mixes aimed at fostering transformative change of socio-technical systems to achieve societal goals. There are currently various evaluation frameworks for TIP, yet these frameworks almost exclusively target at the evaluation of delimited projects, policies, or programmes (Ghosh et al., 2021; Haddad & Bergek;
2023; Boni et al., 2023). This is problematic in cases where TIP consists of a cascade of multi-level multi-sector policy interventions, i.e. policy conducted at various levels and in different parts of government. Policy mechanisms as well as their outcomes interact, which makes it hard to establish additionality of TIP if the evaluation focuses on one program. Haddad & Bergek (2023) dismiss evaluations including “full system analysis” on grounds of time and costs, yet we would argue that the main barrier to such an evaluation currently is the lack of a system dynamics approach that effectively scopes and integrates policy interventions.

Closer alignment with Transition Studies (Kˆhler et al. 2019), characterized by its refined frameworks of system dynamics and ample empirical case studies, seems a viable route to enhance TIP evaluation. However, perhaps related to their origins in the market-focused “long nineties”, main system dynamics frameworks such as the Technological Innovation Systems (TIS) have a fairly implicit understanding of the role of state and (supra)-national

policies (Johnstone & Newell, 2018). Hence, in this study, we explicitly bring “policy acts” back into a system dynamics analysis. We do so by drawing on recent work that allows to establish in more detail different complex form of causality (Geels 2022) which is vital for policy evaluation. We use the notion of “motors” to specify the main interactions patterns fostering socio-technical change, and assess the additionality of policies in these interactions (Suurs & Hekkert, 2009; Bergek & Haddad, 2022).

Our empirical case study of the transition towards electric buses in the Netherlands, can be described as a “success case”. Although it is common in both transition and policy studies to describe failings and learn lessons, an emerging stream in policy sciences stresses the importance of analyzing success cases (Compton & t Hart, 2019). Nuanced analysis of success cases helps to identify and spread policy mechanisms that work. In particular in the domain of sustainability transitions, a sole focus on policy failures and injustices brought about by transitions risks undermining faith (with)in public institutions and their capabilities to address environmental and social challenges. There is thus an urgent need to describe “what works”, not bringing about perfection, but advances towards environmental improvements that take into account social and economic prosperity as well.

We describe the shift from Diesel buses towards battery-electric zero emission (ZE) buses in the Netherlands, a trajectory shaped by transformative policies at multiple levels, as well as private stakeholder involvement. Indicative for the progress is the high adoption rate combined with industrial development. As of 2023, almost all new buses are electric, as well as about 27% of the stock of public transport buses in The Netherlands is electric, making it the leading country in Europe (Ruiz, 2023). In addition, new entrant Ebusco and incumbent VDL Bus & Coach, two of the largest European electric bus manufacturers with 340 and 1728 FTE respectively, are located in the Netherlands. Moreover, public bus services are more often used by people from low and middle income groups. A significant criticism on the transition towards electric mobility so far has been that it has overwhelmingly benefitted wealthy people able to buy expensive electric vehicles (Sovacool, 2019). Subsidizing the development of electric mobility within public transport represents a qualitatively different transition pathway that results in more equitable outcomes, and prevents spending large amounts of public money on the already-wealthy.The success case offers the opportunity to derive lessons that may further the development of TIP research and practice, in particular when it comes to highlighting what mechanisms were most influential for driving change. Taking a longitudinal socio-technical transition perspective, this research provides an original analysis of the role of multi-level policies on complex transition dynamics over time (Alkemade & De Coninck, 2021). Concrete behavioural additionality of policies is determined by tracing changes in stakeholder behaviour due to policy interference (Bergek & Haddad, 2022). Additionally, we investigate how policymakers on different administrative levels (regional, national, and European) addressed several typical challenges related to TIP (Haddad et al., 2022).

First, a database of relevant policies on different levels is constructed, as well as an overview of the procurement process of ZE-buses in the Netherlands. Afterwards, with an interview structure based on a combination of a TIP-evaluation framework from Bergek & Haddad (2022) and practical TIP challenges for policymaking from Haddad et al. (2022), 15 interviews with prominent stakeholders related to the Dutch bus system were conducted to evaluate the ZE-bus transition and the policymaking process associated with it. Using a process-tracing approach (Collier, 2011), the microprocesses accumulating to the ZE-bus transition are first reconstructed into four chronological phases, being: the introduction of competitive tendering and the start of environmental awareness (1992-2010), institutional work and experimentation (2010-2016), administrative agreement and ZE-bus implementation (2016-2020), and the scale-up (2020-2023). After process reconstruction, these elaborate transition micro-processes are sequenced into four “innovation motors” (Suurs, 2009), playing an instrumental role in propelling the ZE-bus transition.

A first innovation motor is the institutional influence by change agents, i.e. the “Foundation Zeroemission Bus” increasing the legitimacy of ZE-buses. This foundation was initiated by a mobility consultancy and a commercial bank, and co-financed by Dutch provinces, national, and European funds. It developed a Total Cost of Ownership model to prove that electric buses could financially compete with existing fossil fuel buses, creating an initial market. In addition, the foundation played a key role in framing and redefinition of values and norms within Public Transport Authorities (PTAs). It also reconfigured political networks, thereby enabling later policy changes.

The second innovation motor is stakeholder alignment through the Administrative Agreement ‘Zero Emission Bus’. The signing of the Administrative Agreement between regional PTAs and the national ministry of infrastructure in 2016 was the strongest articulation of visions and expectations in the electric bus transition. This agreement underlined the shared governmental vision to introduce 100% ZE-buses by 2025, and have a 100% ZE-bus fleet by 2030. Due to the collective agreement, and resulting market perspectives, PTAs built trust with stakeholders and aligned them with a clear transition agenda.

The third innovation motor is reflexive policy learning by Public Transport Authorities. By building and sharing knowledge on tender specifications in new bus concessions and adjusting specifications accordingly, regional authorities built the capacity to translate national transition goals into regional tender demands facilitating innovation. Examples include the right to change bus types during concessions, or a 'handover arrangement' to ensure compensation for investments if the durable ZEbuses would last longer.

The final innovation motor is the organisational flexibility and development of industrial actors, pertaining to the entry or reorientation of Dutch bus manufacturers on the global bus market due to prospective commercial opportunities, and the reorganisation of Public Transport Operators to prepare for large scale ZE-bus adoption. Deriving from the innovation motors, behavioural additionality of policy, and the TIP-challenges, three main policy lessons to positively influence transitions are constructed. These lessons are broadly defined as: identifying and supporting institutional entrepreneurs in preparation for a transition; the formulation of clear, measurable, and trustworthy goals aligning stakeholders towards a common goal; and participating in reflexive policy learning on a regional level during the transition to remove inhibiting policy dependencies. As these critical factors are only partially covered by TIP imperatives covered in the literature, our findings contribute to attempts to make TIP more handson as well as evidence-based. When it comes to the policy evaluation, it has been argued that the emergence of a new policy paradigm, such as TIP for fostering socio-technical transitions, inherently also requires a new evaluation paradigm (Rohracher et al., 2023; Baarslag et al., 2024). Our study draws attention to one particular feature worthy of reconsideration: the entry point from which an evaluation analysis starts out. In traditional evaluations of 'frame 1 and 2' innovation policies, focused on measuring innovation output and impact, the entry point is a policy instrument or program and the 'treatment' it provides (Santos & Coad, 2023). However, as TIP aims to address societal challenges via transforming sociotechnical systems, it becomes critical to understand policy as one of many elements driving change through sequential and configurational causal mechanisms (Geels, 2022). We therefore proposed an evaluation approach that combines transition meta-theories (to conceptually grasp system change) and configurational process tracing (as an empirical methodology) for assessing the system effects of TIP.

The contribution we make here sits at the interface of at least two ongoing academic debates. One concerns the literature on (technological) innovation systems, which still grapples with understanding the role of policy in system dynamics (Raven & Walrave, 2020; Gong & Hansen, 2023). While there is a lively debate on identifying intervention points for driving system change (e.g. Kanger, 2020), less is known about how to determine the influence of such policies (Turnheim et al., 2015). At the same time, scholars concerned with evaluation have acknowledged the demand for evaluation approaches suitable for understanding policy influence in relation systemic evaluation perspectives (Haddad & Bergek, 2023; Wise & Arnold, 2022). So far little efforts have been taken in that direction, perhaps due to research often being conducted in the context of commissioned evaluation studies innately tied to a specific policy instrument or programme. In as far existing studies do aim to empirically assess the systemic impacts of policy, this often involves measuring systemic changes – e.g. based on TIS functions – without specifying which system this would concern (e.g. Janssen, 2019). The approach we have suggested addresses this limitation by starting out with first defining the system, and then asking what role policy placed in transforming it.

Francesco Cappellano, Francesco Molica
WSB University, Poland
European Commission, Joint Research Centre (JRC), Belgium

EVALUATING REGIONAL SMART SPECIALIZATION AGAINST THE MISSION-ORIENTED APPROACH

ABSTRACT

This paper aims at evaluating regional innovation policies – particularly those embedded in the Smart Specialization (S3) rationale - against the transformative Mission-oriented approach. The emerging paradigm of 'transformative' or 'mission-oriented' innovation policy, which addresses broad societal challenges, is subject to a growing body of research. Yet the academic discussion within this strand has largely taken a theoretical perspective, while less attention has been devoted to how this new framing can be translated into concrete policymaking (Bergek et al., 2023; Magro and Wilson, 2019). With the view to address this gap, one specific question that requires scrutiny is the extent to which established policy frameworks can accommodate a mission-oriented dimension. The paper contributes to tackling this question by focusing on smart specialization strategies (S3). Smart specialisation is currently the most widespread approach regarding regional innovation in Europe, despite having been implemented to very different fortunes (see e.g.: Capello and Kroll, 2016; Gianelle et al., 2019; Kroll, 2017). It appears therefore critical to investigate how S3s can effectively bear a transformative change by integrating a mission-oriented approach given their large diffusion.

The paper is divided into two sections. The first one defines a comprehensive theoretical framework articulating the main dimensions of the mission-oriented approach based upon extant literature. The novelty of the framework is that it is designed in such a way to cast light on how S3s can be reconfigured in response to societal challenges (SC)s. It takes account of the specific features and limitations of the smart specialization concept also from a policy practice perspective.

The second part analyses three case studies, i.e. three S3s with an explicit transformative or missionoriented dimension (Catalonia, Northern Netherlands, Czech Republic), against the framework. Such analysis aims to provide evidence and insights as to the concrete challenges of operationalizing the mission-oriented approach into the S3 (e.g. delivering experimental governance; monitoring of directionality; etc.). In so doing, it brings important evidence in relation to how transformative policies can be translated into concrete policy-making processes and what are the main obstacles.

In order to compare and assess Smart Specialisation Strategies (hereinafter S3), against the MOA, this paper hinges on a conceptual framework based on a selection of notions which received attention from the scholars as they define the Mission-oriented approach. The analytical framework spans four broad dimensions discussed below.

This research aims to investigate how S3s can effectively bear a transformative change by integrating a mission-oriented approach given their large diffusion. To do so, the authors engage with different notions which substantiate the concept of the MOA to evaluate S3 on a theoretical basis as well as providing instances from practices, debated in the next section.

The S3 introduced, or at least reinforced, the emphasis on the need to prioritize certain policy domains over others through the Entrepreneurial Discovery Process (EDP): a policy practice to collectively assess the economic areas where the region has its own competitive advantages (Foray et al., 2011). The prioritization and the result-oriented approach introduced by the S3 hints at what is called policy directionality defined as "promoting innovations that contribute to a particular direction of transformative change" (Parks, 2022).

The innovation policy's directionality gained salience (Weber & Rohracher, 2012) informing the whole discourse on modern innovation policies which nurtured a multitude of new policy approaches (see Haddad et al., 2022 for a deeper description). Among those, the Mission-oriented approach - hereinafter MOA - (Mazzucato, 2018) spearheaded the policy landscape incepting the launch of several mission-oriented policies at national and international levels (see the EU Missions). Such approach gained momentum on the idea to align actors and orient resources, strategies toward selected priorities (Janssen et al., 2021).

A policy targeting SC – either in S3 outlet or in other policy frameworks – should show directionality and consistency across the whole categories of policy elements.

On the other hand, the place-based extent of S3 relates to the context-specific policy outcomes which should be prompted through a locally run EDP, avoiding a "me too effect" which was based on adapting strategies produced elsewhere under the "one-size-fits-all" approach (Foray et al., 2011). Scholars argued that a place-based contextualisation of SC in both their framing and solution selection processes is essential to reap benefit of local market opportunities and gain public legitimacy on SC-oriented policy activities (Wanzenbˆck, et al., 2020; Flanagan, Uyarra & Wanzenbˆck, 2023). The surgence of what is called micro missions (Henderson, Morgan & Delbridge, 2024) or local wins reflect an urban or local dimension of MOA based on very place-specific conditions.

Research has shown the different capacities' endowments of European regions concerning administrative (Mendez & Bachtler, 2022), institutional (Farole et al., 2011), governance (Charron et al., 2014), technological (Balland et al., 2019, Bachtrˆgler-Unger et al., 2023) and innovative skills (Cappellano et al., 2022) that underscore the salience of context-specific synergy policies. This generates significant differences in expenditure efficiency (see Bachtrˆgler, Fratesi & Perucca, 2020). In the S3 operational process, public stakeholders are asked to possess a diverse set of capacities. Connective capacities should be mobilized to connect actors, organizations, and networks (Gieske et al., 2016) by determining the policy directionality through an agenda setting process (Kulman & Rip, 2018). This would mitigate the risk of falling into a "directionality failure" (Weber & Rohracher, 2012). Furthermore, collective intelligence practices focus on the analysis of risks and opportunities (Kitagawa & Vidmar, 2023). In fact, a strategic intelligence should be adopted to assess the local opportunities and threats in respect to SCs, recognize key assets in terms of knowledge, resources, and ongoing policy dynamisms (Cappellano et al., 2022; Bugge et al., 2022; Bours et al., 2022), negotiate a policy agenda (Flanagan et al., 2023), prioritize strategic orientation, promote policy coordination, and facilitate policy implementation (OECD, 2021; Pontikakis et al., 2022).

Those capacities refer to future-looking policy approaches such as Foresight which indeed show commonalities with both S3 and MOA as both are aimed to generate results in the future. The authors acknowledge that there is a common forward-looking perspective in both S3 practice and MOA. However, there are sheer differences among them: S3 focus is dominated by the technological and entrepreneurial scope. S3 promotes exploitation of existing technologies through a sustaining change perspective. Hence, it was argued that it promotes path-dependent development (Hassink & Gong, 2019). Instead, the future-looking perspective, at the heart of MOA, promotes innovation toward a transformative change (OECD, 2021). Both in the theoretical understanding (Mazzucato, 2018) and in the EU missions, future-looking ambitious results are set to tackle SCs, while promoting a societal transformation.

The concept of smart specialization assumes a shift in the governance of R&I away from the topdown models of the past (McCann & Ortega-ArgilÈs, 2013). The S3 approach sees in participatory and bottom-up processes a critical factor to chart the right path towards a more competitive specialization. In a specular way, there is consensus in the literature on MOA around the involvement of a wide spectrum of stakeholders in the design and implementation of missions (Mazzucato 2018; Wanzenbˆck et al. 2020; Janssen et al. 2021; Larrue 2021; Jannsen et al, 2023; Wiarda et al., 2023). The main reason is that, given the contested and complex nature of SCs and their solutions, missions demand a great deal of coordination and legitimacy across many actors.

Experimentalism is on paper a key dimension in both MOA (Mazzucato, 2018) and S3 (Radosevic et al., 2017). However, the scope for experimentalism in the S3 has been to date severely restricted by the rigid, audit-driven, and ultimately cost-efficient, nature of cohesion policy rules (Gianelle et al., 2020; Rauhut & Humer, 2020; Molica, 2024), the disproportionate emphasis on design, lack of policy capacity (Laranja, 2020), institutional arrangements (Morisson & Doussineau, 2019). The latter two problems are a common issue in MOA. Additionally, the need to set clear objectives in the MOA can clash with the "muddling-through" approach underlying the experimental policy-making whereas the emphasis on autonomy raise issues in terms of institutional accountability (Radosevic et al., 2023)

The selection of policy instruments is also paramount in innovation policies. Foray (2022) argues that a bad engineering of instruments can compromise even of the most well-thought innovation policies. This is even truer in relation to the design of policy mixes, which, in the context of S3, considering its operationalisation in multi-level settings and pre-existing governance, policy and institutional arrangements, can turn out to be a very complex exercise implying trade-offs and tensions (Reid & Maroulis, 2018; Nauwelaers et al., 2014). This dimension has attracted less attention from scholars than other dimensions of S3.

However, anecdotal evidence seems to show that the use of policy mixes has been largely limited for S3s: in fact, in spite the initial ambition to act as overarching policy frameworks, many strategies have ended up guiding only the R&I investment of ESIF funds (Pontikakis et al., 2022). Even in this case, Gianelle et al. (2019) find that in a non-negligible number of cases there is little alignment between the strategies and policy interventions under ESIF whilst D'Adda et al. (2021) finds scant evidence that S3s have re-shaped ESIF funding decisions.

The complex and cross-sectoral nature of challenges addressed in the MOA requires a strong integration across various instruments (Larrue, 2021). This idea dates back to first works referring to mission-oriented policies such as in Freeman (1996). Mazzucato stresses the importance of using a wide array of financial instruments, including more innovative ones such as innovative public procurement and challenges prices, to cater for different funding needs and risk profiles. Moreover, given the systemic perspective of missions, supply-pulls instruments need to be integrated by demand-side ones (Tˆdtling et al., 2021).

The aim of this study is to develop an analytical framework for evaluating and monitoring how the MOA has been adopted in regional S3s. The framework is based on the different dimensions discussed above by providing qualitative guidelines for assessing S3 documents. The feasibility of the framework will be tested with a number of case study regions from different EU countries.

This exercise is expected to generate a few contributions to the theoretical debate while also providing practical policy recommendations. Firstly, the conceptual framework helps to frame and unpack the term "policy directionality" into practical terms contributing to the few studies conducted to date (see Bergek et al., 2023). The paper enriches the discussion around the pivotal topics of actors' – namely, institutions - capacities reflecting on the place-specific conditions for policy implementation. Finally, this analysis foregrounds advice for policymakers at the EU level concerning the adaptation of current S3 into more modern Challenge-oriented approaches. A sound and evidence-based approach might prevent "reinventing the wheel" exercises, forcing regional institutions to apply a complete newly approach rather than editing the ongoing practices.

Florentine Frantz, Tobias Dudenboste, Philipp Aiginger-Evangelisti
Technopolis, Austria
Technopolis, Austria
FFG , Austria

CHALLENGES OF ACCOMPANYING CYCLES OF LEARNING

ABSTRACT

The FFG initiative “Expedition Zukunft” (engl. Expedition Future) aims to promote disruptive and radical innovations that target the disruption of markets, the solution of complex societal, economic, or ecological problems, or radical technological leaps. The programme is highly ambitious, as the funded innovation projects should also pursue the goal of making all of our lives noticeably and sustainably “better”. The initiative is structured along four project lines (#Start, #Innovation, #Science and #Challenges), focusing on early-stage innovation projects or knowledge transfer between science and industry. While #Start, #Innovation and #Science are open-topic research funding programmes, the #Challenges are looking for interdisciplinary teams of problem solvers responding to challenges for the public sector as well as on “Water and Soil”. In the future, a selection of the projects funded in these programme lines will be supported in internationalisation and scaling in a fifth programme line #Scaling. The programme thus spans the entire spectrum from brainstorming and cooperation to the public sector.

Central to this is the support provided by the Expedition Guides (#Explore) from the programme team, who are intended to provide funding recipients with regular, structured and needs-based orientation in the Austrian RTI funding system. Conceptually, the selected projects are to receive holistic and long-term support, i.e. monetarily, through consulting services from external experts on topics such as business models and advice from the FFG (e.g. on further funding opportunities). As such, the Expedition Guides are intended to support beyond the duration of the funded project on their way through the innovation system to commercialisation. #Explore represents also a conceptual innovation of Expedition Zukunft itself, applying an active agency approach that is also increasing the FFG's knowledge on the development and dynamics of innovation projects. Therefore, Expedition Zukunft also can be understood as a wide-ranging programme that is intended to facilitate organisational learning on various levels.

Technopolis accompanies the Expedition Zukunft team in the framework of an accompanying evaluation. The aim of the evaluation is to identify potential for improvement in communication, funding conditions, selection procedures and support for funding recipients and to develop recommendations for the further development of the programme. Methodologically, we rely on workshops in various formats with the programme team and other representatives of the FFG on the programme concept, programme hypotheses and central programme activities, feedback workshops with applicants, a critical analysis of central programme documents and online texts with regard to their suitability for programme communication, and accompanying observation

of the application process. In addition, we analysed programme data on target group achievement.

In this contribution, we reflect on the challenges of evaluating a complex programme that constantly develops further quickly and significantly. We consider both the experience of the evaluators and the programme team of the implementing agency of the FFG. As the project is still on-going, the outlined reflections are preliminary and will be expanded/deepened until the conference. These reflections are further informed by FFG's internal considerations and deliberations regarding the usefulness of different forms and modes of evaluation towards learning and programme development. Against this backdrop, FFG has decided to request an accompanying evaluation with the intention that identified challenges represent valuable learning opportunities allowing immediate responses. This is especially important for new formats and instruments, where failure and learning are expected. The programme team is also not in a position to wait for the results of a summative ex-post evaluation that also seems to be associated with a higher risk, as identified concerns cannot be corrected as quickly which might threaten further budgets.

Challenges of Accompanying Cycles of Learning: Expedition Zukunft is a pilot initiative where some programme characteristics had to be developed during its implementation. While this was done with high attention to detail and best possible planning at that time, experiences made during programme implementation and evidence collected through the evaluation revealed a constant need to further develop definitions (e.g. "radical", "disruptive", "groundbreaking" innovation), selection criteria and processes, communication tools and their content (homepage, programme documentation, FFG-e-call) as well as adapting FFG internal tools like instruments and guidelines). For our conference contribution, we characterize these continuous adaption processes as Cycles of Learning. We will illustrate the dynamics and challenges for the Cycles of Learning in the collaboration of FFG and Technopolis as well as the Cycles of Learning for the implementation of the programme lines, with a specific focus on #Explore. Structuring our reflection of the Cycles of Learning we will map out the temporal, empirical, and practical challenges of accompanying these processes.

- Temporal Challenges: First, we will reflect on the challenges arising from the high-paced development of the initiative, services, and calls, the time-lag of interventions and observable changes, the rhythms of calls, meetings and workshops, as well as the challenge to match reporting timespans and moments when feedback is most useful. Carving out the temporal challenges, we also highlight the concrete changes of evaluation project processes we collectively developed in response.

- Empirical Challenges: Second, we will describe the empirical challenges that arise in the accompanying evaluation, which cannot observe and accompany all steps and details equally. Implementing four programme lines with independent selection mechanisms and standards meant that the number of important decision-making points and procedures also multiplied. Moreover, the high granularity and the character of a pilot programme meant that at the beginning, possible number of data points were limited and differed between programme lines (e.g. from more than 60 applications to a handful, depending on programme line and call). Many decision-making points were observable for the evaluators, but others were not. The focus of attention is thus put on details and demands of where the FFG requires most support and values an external perspective the most.
- Practical Challenges: Third, we map out a set of practical challenges that arise with the demand for the evaluation as contributing to the legitimisation of the programme, i.e. towards the main funding source (the Nationalstiftung für Forschung, Technologie und Entwicklung - FZÖ) and relevant ministries, but also to enable learning among the programme management. For both functions, different tools, in particular for knowledge transfer and documentation, had to be used.

Emily Lalonde, Matthias Egger
Utrecht University / The Ocean Cleanup, Netherlands

EVALUATING THE TRANSFORMATIVE OUTCOMES OF THE OCEAN CLEANUP'S RIVER PROJECTS

ABSTRACT

Plastic pollution poses a significant threat to aquatic ecosystems and human well-being. Tackling this challenge demands transformation of the existing socio-technical system. Such transformations typically rely on the cumulation of a broad range of activities that, due to synergies amongst them, allow for overcoming the inertia of the existing system. Driving transformative change on plastic pollution are different solution-centered initiatives, which span macro-level interventions like international missions (e.g., Plastics Treaty) to entrepreneurial engagement initiatives undertaken by organizations deploying innovative technological solutions, like ones set by

The Ocean Cleanup which aim to tackle 90% of ocean plastic pollution by 2040 using technological solutions. The Ocean Cleanup is an ambitious nonprofit organization focused on removing legacy plastic from oceans and, more recently, rivers, setting an ambitious mission of 'preventing 80% of floating plastic being emitted through the 1000 most polluting rivers by 2040, by deploying Interceptor solutions'. Such missions demand change across all dimensions (technology, actors, institutions) of the sociotechnical system. The Ocean Cleanup has directly begun to tackle this change. First, by using technological innovation to address ocean plastic pollution. Second, by extending its approach to rivers, tackling ways to prevent plastics from ending up in the ocean in the first place. The extent to which The Ocean Cleanup is achieving transformative change through its activities remains a key organizational question, one that heavily influences its ability to scale and achieve its mission. This question is answered through evaluation.

Concepts in Transformation Innovation Policy (TIP) help understand river deployments, and their transformational success, and provide a basis for evaluation. To evaluate the transformational success of The Ocean Cleanup's River Department, its river mission is conceptualized as a TIP. The goal of the TIP is to eliminate riverine plastic emissions into the ocean; transforming rivers from those that emit plastic, to ones that do not. This connection to evaluating transformative policies relates strongly to the first thematic area of the REvaluation Conference 2024 – The evaluation of transformation policies and dynamics in socio-technical systems.

Elaborating more on TIP evaluation, most existing sustainability transition frameworks, like Quantitative Systems Modelling and Socio-technical Transition Analysis, focus on the macro- or mesolevel of socio-technical change and provide little direction in how to assess the input of more localized efforts. The Ocean Cleanup's river deployments are localized efforts, therefore, evaluation at this level is necessary. Experimentation, drawing from the Multi-Level Perspective (MLP) framework, offers an approach to evaluate local efforts by exploring concepts like niche, regime, and socio-technical landscapes.

Experimental Policy Engagements (EPEs) represent varying degrees of experimentation, including projects focused on one river deployment, broader programs connecting multiple initiatives within a region, and policy mixes involving alliances with organizations addressing systemic issues of plastic pollution. The theory suggests that EPEs become more transformational when multiple initiatives and other dimensions of socio-technical systems are employed (so programs and policy mixes).

EPEs, however, cannot be used to evaluate the transformational success of policies as, alone, they do not result in transformative innovation. More broadly, transformative change often encounters problems related to the absence of directionality, misdirection of demand expression, insufficient policy coordination, and lack of reflexivity. To address challenges in orienting change efforts and the shortcomings of EPEs alone, transformative outcomes (TOs) are useful.

Transformative outcomes are (12) processes or solutions that result in deeper transformations, leading to changes in the rules that influence actors' (individuals, groups, organizations) behavior. Examining the prevalence and determinants of TOs through this study facilitates cross-learning and identifies factors for successful replication as The Ocean Cleanup scales its operations toward its goal of making all oceans plastic-free. EPEs accomplish TOs via continuous observation, evaluation, trial, and reflexivity.

Speaking of the novelty of this evaluation approach, there are limited papers on this topic and a demonstrated need to test TOs in different contexts to understand what outcomes apply. Additionally, there is a lack of knowledge on how to increase the quality of outcomes and how to extend ongoing actions to accomplish more transformation. Additionally, evaluating TIPs initiated by a non-state actor is another novelty of this study, as current debate focuses on state-led missions.

This study's evaluation of The Ocean Cleanup's mission or TIP employs a deductive, qualitative, and formative approach. Deductively, this study applies transition theories (EPEs and TOs) using the unique, single case of The Ocean Cleanup from which four embedded subcases (EPEs of different locations, levels, and speeds) provide a better understanding of variation and highlight the pitfalls of the existing theory. The formative use of TOs in a comparative setting allows for a nuanced understanding of the contextual factors driving (or hindering) transformative outcomes across diverse implementations of similar experiments, such as The Ocean Cleanup's river deployments. This approach facilitates the iterative refinement of policy initiatives by highlighting the relationships between experiments, contexts, and outcomes, ultimately enhancing understanding of systemic transition processes and guiding future transformative endeavors. Further other qualitative and formative methods such as Theories of Change and outcome harvesting operationalize concepts from the transitions literature and aid in evaluating the success of the TIP.

The analysis reveals nuanced relationships between place-specific contextual factors (e.g., community involvement), EPE-level dynamics, and temporal evolution in determining the transformative success of river deployments. Higher-level EPEs that prioritize community involvement and perspectives in the experiments tend to result in deeper transformations, while time allows for the emergence of broader and more profound changes.

The results also highlight the significance of ongoing monitoring and adaptation in ensuring the effectiveness of river deployment initiatives. By considering socio-technical contexts and embracing a flexible approach, organizations can optimize their strategies and accelerate progress toward addressing plastic pollution in rivers and oceans.

Additionally, the study challenges conventional notions by highlighting the role of non-state actors, like The Ocean Cleanup, in driving transformative change beyond local-level impacts. By engaging diverse actor networks and leveraging community-centered activities, organizations can enhance their impact on transformational environmental sustainability goals.

In conclusion, this research on The Ocean Cleanup's mission to combat plastic pollution in rivers offers valuable insights into the evaluation of transformation policies and dynamics in socio-technical systems, a key thematic area of the REvaluation Conference 2024. By conceptualizing river deployments as EPEs within the framework of TIP, the need for effective approaches to monitor and evaluate research and innovation policies aimed at driving socio-technical transformation is addressed. Overall, the research contributes to the advancement of evaluation practices in supporting sociotechnical transformation towards a more sustainable future, offering actionable insights for policymakers, evaluators, and practitioners striving to promote environmental sustainability and drive socio-technical change.

Dr.in Petra Biberhofer, Mag.a Aliette Dörflinger, Prof. Dr. Christian Pohl, Dr. Uwe von Ahsen
Austrian Science Fund, Austria
Dörflinger Consulting & Facilitation, Austria
Transdisciplinarity Lab, ETH Zürich, Switzerland
Austrian Science Fund, Austria

REFLECTIONS ON THE #CONNECTINGMINDS PROGRAMME OF THE AUSTRIAN SCIENCE FUND FWF - KEY LEARNINGS IN COMPETENCE BUILDING ON TRANSDISCIPLINARITY

ABSTRACT

There is a certain history of transdisciplinary funding programmes in Austria, which have been implemented by various institutions over the last 25 years. These have often been programmes with a thematic focus or project funding. The spectrum ranges from programmes such as Cultural Landscape Research with a focus on the humanities and social sciences in the 1990s to the current Earth System Science Programme of the Austrian Academy of Sciences or the Cancer Mission Lab of the Open Innovation in Science Centre of the Ludwig Boltzmann Society. However, it was a special step in the research policy landscape when the Austrian Science Fund FWF 2020 included the transdisciplinary funding programme #ConnectingMinds (CM) in its funding portfolio. The FWF is Austria's leading organisation for open-topic funding of basic research as well as artistic and scientific research at a high international level. Each year, the FWF Board decides on the funding of more than 2,700 research proposals, which are reviewed by around 5,000 international scientific experts. In 2022, its funding budget, allocated by the Federal Ministry of Education, Science and Research, will amount to approximately €273 million. A new three-year agreement has increased the FWF's budget to around €350 million between 2024 and 2026. The FWF's funding approach is strictly "bottom-up", i.e. the research topics are proposed by the researchers in all areas and all programmes and there are rarely thematic calls for proposals.

In this contribution, which offers a transnational perspective from the perspective of the programme management (FWF) and the jury chair (ETH Zurich), the aim is to share experiences and critical reflections on the development of the transdisciplinary programme with other funding bodies, universities and the TD community. The contribution spans the arc from the development of the programme (2018-2020) and the considerations behind it to the experiences gained from its implementation in two calls (2020 and 2023/24). The contribution focuses on four key learnings: (1) the need for active programme management with regard to competence building in transdisciplinary research within the funding organisation and externally at universities and colleges; (2) the ongoing involvement of TD experts in programme design and in the review process; (3) the relevance of an open and adaptive institution including decision-makers who are committed to the necessary competence and capacity building and thus stand behind the programme; (4) the orientation towards the TD research community and involved practice actors as a (new) target group.

When the FWF announced the CM programme for the first time in 2020 with a total volume of 4.6 million euros from the National Foundation for Research, Technology and Development, it sent a strong signal in the direction of transdisciplinary research. In particular, because the programme has very high scientific standards in the tradition of the FWF and the high project funding and relatively long project duration (approx. €1 million per project for 5 years) enable the active involvement of so-called practice actors in the research project from the very beginning. The 2nd call for proposals is currently running (spring 2024).

With the #CM programme, the FWF encourages researchers to involve non-scientific stakeholders in research projects. In addition to 2 to 5 internationally outstanding researchers at one or more Austrian research institutions, the programme is aimed at practitioners, i.e. organisations/associations/institutions that are affected by a complex social problem, feel affected by it or have an influence on it, such as representatives of NPOs/NGOs, associations, public administration, companies, health or educational institutions - private individuals are decidedly excluded. Funding is available for teams that combine scientific and social knowledge in order to meet the social, technological, ecological and economic challenges that lie ahead. The focus is on transdisciplinarity (TD), innovative research approaches and the support of social commitment as well as the promotion of collective learning spaces.

Compared to previous initiatives to promote transdisciplinary research in Austria, other programmes have been consistently thematically oriented. One of the FWF's paradigms is the open-topic and open-result design of funding programmes, with the basic assumption that high-quality and excellent research will - sooner or later - have an impact on the economy and society. The FWF remains true to this scheme in the design of the CM programme; it only insists on a socially relevant problem orientation, the result remains open, above all because the project is fully funded and does not necessarily have a bias or even a result orientation due to co-financing from participating stakeholders. In addition, it helps the FWF to communicate the importance of funding basic research to the public through the very practical project topics. In this respect, CM is the first programme in Austria to be set up independently of disciplines, open to all topics (at least in the first call) and politically independent. The process character of TDR is reflected in the process orientation and structure of the programme, which is divided into two stages. The funding of a workshop in stage 1 (FWF: 2023) provides a strong impetus towards networking and sustainable team development. This is essential, especially for transdisciplinary projects, so that the problem framing phase can take place together with practitioners, i.e. relevant nterests can be included in the formulation of the research project right from the start. The planning and discussion of financial resources and the involvement of practice stakeholders is also necessary in stage 1 and should be incorporated into the process of problem identification and the deepening of the transdisciplinary research process in stage 2 (application for the CM project). In the review process, TD expertise is combined via the CM jury and a standard peer review process. The demand for scientific excellence together with a wellfounded transdisciplinary approach is both decisive and challenging for quality assurance.

Key learnings:

- In the case of #ConnectingMinds, the success of the design, implementation and institutional sustainability of such a TDR programme was strongly supported by a top-down approach.
- An institution and staff that are open to learning and innovation are important success factors. Awareness-raising and educational work at the FWF was necessary to understand the relevance and added value of TDR in the context of basic research funding.
- Active programme management is of great importance in order to keep a close eye on this research approach and all those involved. Accompanying measures are certainly crucial for the analysis and further consolidation of the programme. This also means that sufficient (time) resources must be made available to the programme management.
- The involvement of TD experts from the very beginning, both in the design and in the evaluation phase, will ensure high quality throughout the different phases of the programme.
- The community of transdisciplinary researchers is small and spread across different thematic areas. Transdisciplinary researchers work on issues such as sustainability, gender equality, health, new technologies or transdisciplinarity in the arts. They often have deep expertise in their field. It takes a lot of effort to familiarise oneself with a new subject area, especially if transdisciplinarity has a different role in that area, e.g. as a problem solver or as research that critically questions (problematises) the prevailing problems and solutions. Thematically narrow calls for proposals, which at best still specify a view of the problem, are therefore only attractive to a small part of a small community.

- The willingness of researchers in the humanities, social sciences and cultural studies (SSH) to engage in research-driven contributions addressed by EU missions varies - not least because the EU missions still address the broad inter- and transdisciplinarity of research.
- Here it is important to address competence and capacity building for TDR comprehensively through FWF funding as well as through programmes and measures at research institutions. A transnational perspective is not only useful, but indispensable in order to further strengthen and develop the research community in Austria.

This experience with this funding programme in Austria could provide further areas for learning and reflection:

- What is the role of the funding organisation in strengthening or building TDR in Austria? What is the role of research institutions and universities in this community development?
- What is the role of programme management in the context of TDR and how does it differ from PM in "conventional" funding programmes? What areas need more resources? Is more time needed for the target group or not? How much community building and mentoring is involved in PM?
- What is the role of external TD experts? Who is involved in supporting the programme management, the jury and possibly the steering committee? What is the role of TD experts in capacity building? And how can this expertise be built up in the AT, comparable to the academies in Switzerland?
- To what extent has the thematic focus (e.g. EU missions) in the second call of the #CM programme possibly influenced the willingness of the community to engage in research-led contributions?
- What is the role of #CM as an instrument to address politically desired issues (e.g. EU missions) but also to increase the acceptance of research (instrumental benefit of TDR for funders)?
- How can a conflict of objectives between the potential of mission-driven research and the exclusion of the humanities in particular from funding programmes be avoided?

Igor Oliveira, Bianca Cavicchi
Millenium Institute, United States
European Commission , Belgium

ASSESSING R&I POLICY FOR TRANSFORMATIVE CHANGE WITH SYSTEM DYNAMICS

ABSTRACT

A transformative innovation policy (TIP) is vital for balanced growth, competitiveness, and addressing global challenges while aligning with the Sustainable Development Goals (SDGs)4. By embracing holistic innovation that considers economic, social, and environmental aspects, countries can create new economic opportunities, enhance productivity, and tackle pressing issues like poverty and climate change.

Evaluation frameworks for transformative innovation policy are still underdeveloped (European Commission, 2023a). Traditional economic methods and metrics, such as R&D expenditure, GDP, patents, and publications are valuable for assessing economic impacts of research and innovation policy, but insufficient for evaluating transformative change. They primarily measure inputs and outputs rather than the broader, long-term impacts and systemic changes required for transformative change.

Transformative change happens over a long period of time and involves continuous, iterative processes that traditional metrics are not suited to track alone. Furthermore, by considering new metrics that account for social, environmental, and longitudinal economic impacts, we can gain a more comprehensive understanding of R&I policy performance and design policies that do better at fostering transformative change, ultimately supporting sustainable and inclusive growth.

System-based methods can provide a valuable complementary framework to more mainstream methods (European Commission, 2023). In this paper, we aim to:

- Demonstrate the complementary advantages of using system dynamics simulation to enhance the ex-ante impact assessment of R&I policy and programmes for transformative change.
- Provide relevant measurements of socio-economic and environmental impacts.
- Provide insights on trade-offs of policy interventions.

In principle, many of the modelling components necessary to employ the method to assess transformative innovation policy (as defined by Schot & Steinmueller 2018) are present in the SD literature. What are the advantages and added value of modelling R&I policy for transformative change using SD? Are there barriers and challenges that emerge, and how to overcome them?

In this paper, we propose two applications of system dynamics modelling to transformative innovation. In the first one, we present an exploratory SD model (Homer 2014) used to explore impacts of economic growth on environmental performance. Although limited in validation and specific case calibration, this model can serve as a starting point for further development and support stakeholders' understanding of the

impact of innovation policies on transformative change. Sensitivity analyses reveal potential system behaviours, showcasing the utility of these models. In the second one, we perform an ex-ante impact assessment of R&I policy options for the European Framework Programme for R&I, based on the transformative innovation policy paradigm and using system dynamics simulation. We do this by modelling two technological domains, namely the mRNA technology and green hydrogen in the time horizon from 2007 to 2035. This time horizon reflects the Framework Programmes for R&I from FP7 to the end of FP10 to make it relevant for ex-ante impact assessment.

System dynamics (SD) is a computer-based simulation approach for policy analysis in complex settings (Sterman 2018). Its key features are the consideration of feedback and time delays in complex systems, departing from an endogenous perspective, meaning a hypothesis on how current systems (in our case, innovation systems) generate their own dynamic problems that evolve over time (Richardson 2011). SD's policy-inclusive models allow policymakers to simulate different scenarios based on hypothetical policy portfolios and assess their possible outcomes on the chosen dimensions.

SD has been employed to interrogate innovation diffusion (Maier, 1998), technology transitions towards sustainability (Struben 2004), technological innovation systems (Walrave & Raven 2016), regime shifts in the context of resilience studies (Herrera 2017), participatory decision making for sustainability (Videira et al. 2017), sustainable development planning at the national level (Collste et al. 2017, Allen et al. 2024), innovation system policy analysis (Uriona & Grobbelaar 2019), linkages between product design and different phases of adoption (Chavy-Macdonald et al. 2019) as well as policy monitoring & evaluation at the European level (Cunico et al. 2021).

Our point of departure is the long causal chain that separates contemporary research & innovation funding from its sustainable development outcomes, also known as beyond-GDP metrics (Hoekstra 2019). It comprises the outputs of different funding mechanisms that feed into innovation ecosystem capabilities and R&D pipelines of market actors, which trigger diffusion and adoption processes when they release new products. Causal hypotheses are necessary to compute the impacts of such diffusion and adoption processes into sustainable outcomes.

FIGURE 1: CAUSAL CHAIN THAT SEPARATES R&I FUNDING AND CONTEMPORARY BEYOND-GDP DEVELOPMENT METRICS

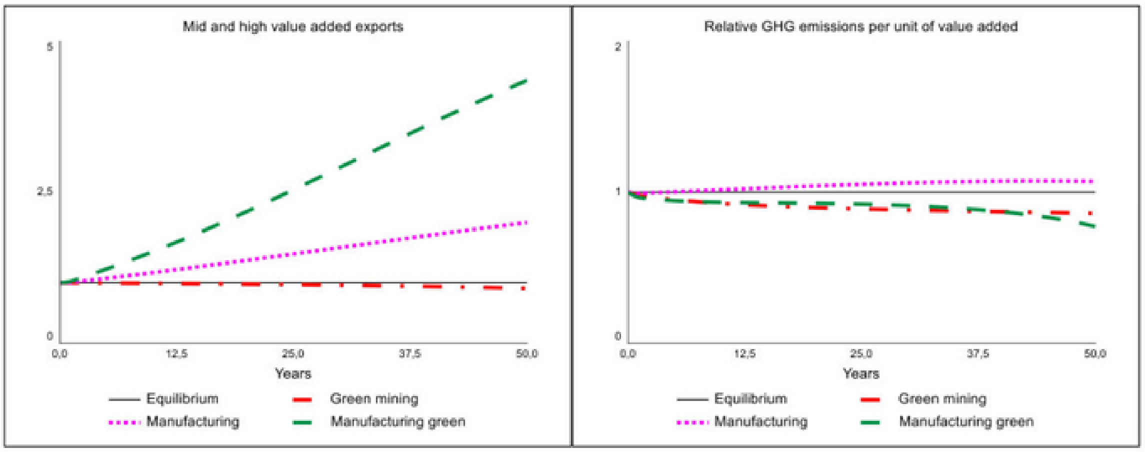
Starting with the most established technological domain, i.e., mRNA technology, we model the feedback relationships across the above-mentioned stages of the innovation process and its sustainability impacts. The R&D pipeline includes R&D funding and project outputs (e.g., publications and patents), corporate R&D and related outputs (i.e., products released on the market). The diffusion and adoption aspect contains production and sales-related variables. Sustainability indicators are mostly based on the SDG framework and span across SDGs 3, 7 and 9.

One of the key strategies to generate confidence in causal models that contain structural uncertainties (as they describe complex phenomena not fully understood by theory) is the use of generic model structures that have been broadly validated in established literature. Therefore, we are employing ageing chains, co-flows and diffusion models (Hines et al. 2011), which are common in representations of dynamic processes with important delays and nonlinearities.

Moreover, whenever an economic phenomenon needs to be represented, we seek to do it via canonical economic models. We have introduced a Cobb-Douglas production function (Cobb & Douglas, 1928) to compute broader effects to be factored in via the computation of economic activity. For example, green hydrogen might make some manufacturing process less resourceintensive, which would create economic growth in these specific sectors. This sector-specific growth might impact the computation of other indicators. In the case of indicators that are calculated in relative terms to GDP, it would make the denominator higher. In the case of more absolute indicators, it might have the opposite effect.

First application: The model presents a simplified Cobb-Douglas function linking medium and high-tech exports with CO2 emissions, demonstrating interconnections between emissions, exports, and production in five economic sectors, i.e., agriculture, manufacturing, extractive industries, non-specialised services, and specialised services. The model includes capital, resource efficiency-oriented capital, and human capital stocks, with total factor productivity influenced by public investment, resource intensity, and inflation. Moreover, the model operationalises the effects of diversity in production capabilities on economic growth, using exports as a proxy for competitiveness and sophistication. SDG indicator 9.4.1 is used to account for all greenhouse gases allowing for reflections on trade-off reflections across economic sectors. The model simulates four scenarios: 'equilibrium,' 'green mining,' 'manufacturing,' and 'manufacturing green.'

FIGURE 1 – SIMULATION RESULTS FOR DIFFERENT SCENARIOS



Results of scenarios simulation highlight trade-offs and synergies between emissions and exports, with the 'manufacturing green' scenario showing potential for synergistic environmental and industrial policies. The study also highlights significant data gaps, emphasizing the need for comprehensive national-level data collection to refine these models.

Second application: The modelling process is now yielding preliminary results on the mRNA technology domain. Our base scenario endogenously replicates the generation of project outputs from European funding as well as the quantities and timings of R&D outputs (in terms of mRNA-based vaccines released to the market) and the adoption process, which in the case of these vaccines is very much public procurement-induced. We have experimented with scenarios that simulate different levels of (grantbased) project funding, but the structural analyses, as well as the simulation of investments in other stages of the innovation process, are still in their early stages. By combining different investment instruments in the model, we expect to reflect on the synergies in trade-offs across transformative innovation policy options, and what portfolios would make sense to maximise the role of Europe in the much-needed sustainability transitions.

SD modelling is proving advantageous to create a bridge in the understanding of how R&D and innovation diffusion and adoption are connected and mutually influence each other. For example, in the case of mRNA vaccines, there has been European funding at several stages, including fundamental research, capital goods technology, process development, drug development and direct payment for production in the form of public procurement. These types of funding contribute to private R&D in different ways, either more upstream or downstream, which mean the delays between the funding and the desired outcome are also different. A quantitative understanding of this dynamics is key to prioritising within TIP-based options, knowing that both the upstream research and the concrete impacts are necessary.

The model can also be used to identify data gaps and reflect on what datasets about project funding, innovation ecosystems, R&D pipelines, innovation diffusion and sustainable development trajectories themselves would allow us to compute the relationship between TIP and sustainable development metrics with a reasonable degree of structural and behavioural validity to allow decision support. Some of the identified data gaps refer to the uptake of project outputs (publications, methods, processes, and patents) by industry, others relate to the nature of these outputs (whether they contribute more to one part of the R&D pipeline or the next). The role of these research projects in the larger landscape of global collaboration within each of the technological domains also needs further investigation. Human capital and tacit knowledge aspects are largely underreported. Some of the non-grant funding mechanisms, such as equity investments, also lack some public data that would allow this type of dynamic analysis.

Michael Dinges, Christiane Kerlen, Jakob Kofler, Kathleen Toepel
AIT, Austria
Kerlen Evaluation, United Kingdom
KMU Forschung Austria , Austria
Kerlen Evaluation, Germany

THE CONTRIBUTION OF R&I PROGRAMMES TO THE GREEN TRANSITION AND THE ENERGY TRANSITION: EVIDENCE FROM INCORPORATING A MULTI-LEVEL-PERSPECTIVE IN SOCIO-TECHNICAL TRANSITIONS RESEARCH

ABSTRACT

The pressing need for a paradigm shift in response to escalating human-induced environmental change has fuelled the quest for a Green Transition in policy, economy, and society. Defining the Green Transition remains an ongoing challenge, as discussions on the scope of sustainability, naturesociety interactions, and diverse actors' roles still shape the discourse. Interdisciplinary perspectives on justice, policy failures, social innovation and wicked problems also highlight the complexity of Green Transition challenges going beyond common policy pursuits for technological advances and reorientation of markets.

In the European Union, the 2019 European Green Deal marked a turning point in the political landscape, taking up instrumental concepts like net-zero and climate targets alongside sectoral policies. However, critiques about the aspired transformation process within the Green Deal, the Green-Growth paradigm, and limits to growth remain unresolved, raising questions about the desired nature of the Green Transition. Growth-critical concepts highlight the constraints on human impact, adding depth to the understanding of the Green Transition. In light of these challenges, the role of the Framework Programmes on Research and Innovation (R&I) becomes crucial. Transitionoriented (R&I) programs can serve as catalysts for innovative solutions, promoting the development of sustainable technologies and innovative practices to navigate the complexities of the Green Transition. Ultimately, they could play a pivotal role in shaping a more resilient and sustainable future.

In this paper, we analyse the role R&I programmes can play in transition processes, studying their contribution to the Green Transition in the evolution from Horizon 2020 (H2020) to Horizon Europe (HE) and comparing it to the contribution to the Energy Transition in the development of the German Energy Research Programme.

The paper takes a transition theory perspective and provides empirical evidence on transition processes that have been induced by the Framework Programmes (FP) and the German Energy Research programme. From the perspective of a policymaker, the paper mainly contributes to the understanding of the contribution of R&I programmes to transition processes. For the research area of R&I policy evaluation and STI policy study, the paper provides novel evidence on the effects of R&I from a transition perspective and contributes to the advancement of R&I policy evaluation frameworks and methods.

The analytical framework for analysing the contribution of the FPs to the Green Transition and the contribution of the Energy Research Programme to the Energy Transition [Energiewende] in Germany rests upon 1) a conceptualisation of the Green Transition in the context of an R&I programme, and 2) an enhanced programme theory approach (cf. Dinges et al. 2022) that combines programme theory (Rogers 2014) with the multi-level perspective of system innovation and the concept of transformative outcomes proposed by Ghosh et al. (2021). The first (1) provides the framework to analyse a) the reorientation of R&I towards net-zero solutions, b) the degree to leverage and scale up existing solutions, c) enabling behavioural change and enhancing capacities, d) preventing harm and repairing damage. The latter (2) is being used to understand better how the Programme's induce impact related to the Green Transition and Energy Transition, respectively.

Methodologically, this research draws upon two evaluation studies that have been carried out independently. Both studies used a mixed-methods evaluation approach, in which participant surveys on transition processes induced by the programmes played a key role. The Green Transition evaluation was part of the back-to-back approach for the ex-post evaluation of Horizon 2020 and the interim evaluation of Horizon Europe on behalf of the European Commission, with a thematic focus on Green Transition aspects and the long-term impact of the Framework Programmes (European Commission 2023).

The evaluation of the Energy Research Programme started in 2021 and is a five-year accompanying evaluation of the measures of the 7th Energy Research Programme (EFP) on behalf of the German Federal Ministry for Economic Affairs and Climate Action. It uses a mixed-methods approach to develop ongoing analyses, reflections and recommendations as a basis for steering and continuous improvement of the programme ("programme learning"), while also contributing to an assessment of its effectiveness and impact of the programme. It fed into the design of the 8th Energy Research Programme, which was launched as a mission-oriented research programme in 2023, with funding starting in 2024.

Green Transition (Dinges et al. 2024, forthcoming): Concerning the contribution of FPs to the Green Transition following the Multi-Level Perspective and the embedded concept of transformative outcomes, most respondents in an online survey indicated that their projects contribute particularly well to the macro-processes of ‘Building and nurturing niches’ and ‘Expanding and mainstreaming niches’. No significant differences were found across the different Societal Challenges or Clusters, and anticipated results from Horizon Europe exceed results from H2020. Within the different transformative outcomes, the e results for ‘Institutionalisation of new strategies and norms relevant to the Green Transition’, results are significantly lower for both programmes, calling into question the boundaries of an R&I programme, and the links between R&I, policy making, and deep learning of system actors. In terms of ‘Opening up and unlocking regimes’, contributions of the FPs are distinctly lower than to the other processes. Difficulties to mainstream the solutions developed through the Framework Programmes and changes in behavioural interactions between old and new fields of the innovation system persist. Furthermore, synergies between the FPs and other funding mechanisms are not being actively developed.

Energy Transition: The respondents' assessments of the macro-processes lead to predominantly critical evaluations of the status of the transformation processes for energy systems in Germany. Based on the initial results of the evaluation (Dinges et al. 2023), it is clear that the activities funded by the 7th EFP have made good contributions to technology development, the promotion of innovations, and the demonstration and application of new solutions in new contexts. The R&I projects focussed on individual technologies contribute in particular to the macro-process of ‘Building and nurturing niches’. Living labs support expanding and mainstreaming niches. The accompanying measures (e.g. energy research networks) successfully facilitate learning and the exchange of experience at the project level as well as raising awareness of new innovative solutions.

Relevant needs for the energy transition that go beyond the traditional focus of a research programme, however, receive little support. From this holistic perspective on the Energy Transition, the results indicate weak contributions to the macro-process of ‘Opening up and unlocking regimes’. Organisational and social innovations, which are essential for the energy transition, have so far been underrepresented in the energy research programme.

A comparison of changes over time will be available for the conference as a second survey has been concluded and is being analysed at present. The positioning of the 7th Energy Research Programme in the context of national and international examples shows that the energy research programme is primarily aimed at the provision of new technologies (supply orientation) but also wants to focus on accelerating the transfer of new knowledge and technologies (demand orientation) and system development (social orientation).

A status quo analysis of R&I interventions for the energy transition in Germany has shown that they are extremely diverse. So far, a supply-oriented approach with little societal orientation has dominated. The fragmentation and duplication of measures and the large number of different funding organisations pose challenges. However, there are learning opportunities and good practices at supranational level that can serve as a model for a stronger systemic orientation of the EFP.

Both evaluations also analysed the involvement of stakeholders in programme planning and the funded projects. The analyses for the Energy Research Programme show that stakeholder groups outside the direct target groups, research institutions and industry are reached to a lesser extent (Dinges et al. 2023). For the FPs, it becomes evident that although the involvement of regulatory authorities and standardisation bodies has improved, stakeholder involvement is still not sufficient in some areas.

Both programme evolutions show that although the transformation orientation has improved in the case of the FP from H2020 to HE, it has reached its limits, particularly in the 'Unlocking Regime' macro-process. The results of the EFP confirm the findings from the evaluation of HE in the level comparison. We assume that there will be little evidence for change in the EFP over time because the design of the programme has not changed fundamentally. The results also underline that the linking and interfaces between different funding programmes are particularly important.

With regard to the funding landscape for the energy transition in Germany, the mission orientation of the energy research programme, which is geared towards overarching goals and challenges, requires that interfaces with other funding programmes should be considered and defined. In order to realise the energy transition, a mixture of various supply and demand-side policy instruments will be necessary to initiate and support processes of system change. The instruments can range from research and innovation funding to regulation and procurement (e.g. green and innovation-orientated procurement, pre-competitive processes), whereby there is a need to experimentally test the interaction of different supply and demand-side policy instruments. Despite the increased ambition of STI policy in recent years, it is clear that knowledge and tools from other policy areas are needed to achieve the green transition. The challenge here is to break down silos and coordinate actions across different government departments. The cases illustrate that while the importance of coordination is clear, long-standing institutional practices hinder its implementation. Thus, without deeper changes in institutional practices, the implementation of transition-oriented R&I programmes may fall short of expectations.

REFERENCES:

- Dinges, M., Coatanroch, G., Kerlen, C., Soloviy, V., Toepel, K., Montalvo Rojo, A. (2024): Navigating the Green Transition. Unveiling the Impact of European Framework Programmes for R&I through a Multi-Level-Perspective in Socio-Technical Transitions Research (Forthcoming)

- Dinges, M., Kerlen, C., et al. (2023): Begleitevaluation der BMWK-Fˆrderung im 7. Energieforschungsprogramm. Erster Zwischenbericht, Stand Juli 2023.
- European Commission, Directorate-General for Research and Innovation, Dinges, M., Coatanroch, G., Evaluation study on the European framework programmes for research and innovation for addressing global challenges and industrial competitiveness – Focus on activities related to the green transition – Final report phase 1 – Horizon 2020, Dinges, M.(editor), Coatanroch, G.(editor), Publications Office of the European Union, 2023, https://data.europa.eu/doi/10.2777/422725
- European Environment Agency (2021), growth without economic growth, Briefing no. 28/2020, ISBN 978-92-9480-320-7
- EEA (2018). Perspectives on transitions to sustainability. European Environmental Agency, Report 25/2017, Luxembourg.
- Leonard, M., Pisani-Ferry, J., Shapiro, J., Tagliapietra, S., & Wolff, G. B. (2021). The geopolitics of the European green deal (No. 04/2021). Bruegel policy contribution.
- Ossewaarde, M., & Ossewaarde-Lowtoo, R. (2020). The EU's green deal: a third alternative to green growth and degrowth?. Sustainability, 12(23), 9825.
- Almeida, D. V., Kolinjivadi, V., Ferrando, T., Roy, B., Herrera, H., GonÁalves, M. V., & Van Hecken, G. (2023). The "Greening" of Empire: The European Green Deal as the EU first agenda. Political Geography, 105, 102925.
- Mastini, R., Kallis, G., & Hickel, J. (2021). A green new deal without growth?. Ecological Economics, 179, 106832.
- Hermoso, V., Carvalho, S. B., Giakoumi, S., Goldsborough, D., Katsanevakis, S., Leontiou, S., ... & Yates, K. L. (2022). The EU Biodiversity Strategy for 2030: Opportunities and challenges on the path towards biodiversity recovery. Environmental Science & Policy, 127, 263-271.
- Geels, F.W., 2006. Multi-Level Perspective on System Innovation: Relevance for Industrial Transformation, in: Olsthoorn, X., Wieczorek, A.J. (Eds.), Understanding Industrial Transformation: Views from Different Disciplines, Environment & Policy. Springer Netherlands, Dordrecht, pp. 163– 186. https://doi.org/10.1007/1-4020-4418-6_9
- Geels, F.W., Sovacool, B.K., Schwanen, T., Sorrell, S., 2017. The Socio-Technical Dynamics of Low Carbon Transitions. Joule 1, 463–479. https://doi.org/10.1016/j.joule.2017.09.018
- Ghosh, B., Kivimaa, P., Ramirez, M., Schot, J., Torrens, J., 2021. Transformative outcomes: assessing and reorienting experimentation with transformative innovation policy. Sci. Public Policy 48, 739–756. https://doi.org/10.1093/scipol/scab045

- Kemp, R., Schot, J., & Hoogma, R. (1998). Regime shifts to sustainability through processes of niche formation: the approach of strategic niche management. Technology analysis & strategic management, 10(2), 175-198
- Rip, A., & Kemp, R. (1998). Technological change. Human choice and climate change, 2(2), 327-399.
- Schot, J. (1998). The usefulness of evolutionary models for explaining innovation. The case of the Netherlands in the nineteenth century. History and Technology, an International Journal, 14(3), 173- 200.

Florence Benoit, Erik Canton, Valentina Di Girolamo, Dario Diodato, Julien Ravet
European Commission, DG Research and Innovation, Belgium
European Commission, DG Research and Innovation, Belgium
European Commission, DG Research and Innovation, Belgium
European Commission, Joint Research Centre, Sevilla, Spain
European Commission, DG Research and Innovation, Belgium

WHY COMPLEXITY IS KEY TO ASSESS EU R&I POLICY

ABSTRACT

In today's economy, knowledge represents a critical resource for long-term economic and sustainable growth (Romer, 1990). Knowledge tends to be highly clustered in larger and more densely populated areas due to the sensitivity of knowledge spillovers to geographic distance. By residing in close geographical proximity, knowledge actors such as firms and individuals can generate a web of social interactions and professional networks while overcoming coordination and incentive problems. This creates an environment that enables the rapid diffusion of ideas and the recombination of various types of capabilities (Storper and Venables, 2004), thereby providing opportunities for the development of new products, technologies and knowledge domains and the creation of a knowledge-based economy. In the process, tacit knowledge plays a crucial role. Unlike explicit knowledge, which can be easily codified and transferred, tacit knowledge builds around personal experiences, skills, and insights that are often difficult to articulate, share or copy. As such, it is viewed as a key source of competitive advantage (Hidalgo and Hausmann, 2009), which is deeply rooted in space (Balland et al., 2020), as it drives the development of unique capabilities and expertise that cannot be easily replicated by competitors.

Understanding these unique local capabilities and the creation thereof is crucial to uncover a territories' specialisation and technological trajectories. From a policy perspective, understanding which regions or countries hold the most valuable type of knowledge and how different innovation systems diversify and develop over time is of key relevance. Such insights may provide directions into how existing capabilities can be leveraged as an asset towards the pursuit of ambitious EU policy objectives, including boosting Europe's long-term competitiveness and improving living standards (European Commission, 2024). Consequently, they are a crucial input for Research and Innovation (R&I) policies that aim to go beyond more traditional innovation approaches, focusing solely on technological innovation and economic growth, towards the creation of a policy framework that is able to address grand societal challenges (Cavicchi et al., 2023).

Such a carefully designed R&I policy is usually technologically specific and focused on relevant knowledge ecosystems. Therefore, it necessitates informed, data-driven policies ó

grounded in sound data analysis and evidence-based decision making. To facilitate such analysis, existing policy and academic literature have put forward various indicators. Nevertheless, many of the traditional indicators rely on R&D expenditure and patent data to determine the innovation performance of countries and regions. Although providing valuable insights, these indicators typically focus on the quantity of knowledge outputs produced, thereby implicitly assuming that all knowledge has the same value and fail to fully capture all nuances of technological progress (Balland and Rigby, 2017).

Against this background, the concepts of knowledge complexity and relatedness have been receiving increased attention in policy analysis and evaluation. Complexity in economic systems refers to the multifaceted nature of economic and knowledge activities and their interdependencies, and captures the structural diversity and the division of knowledge within regions or countries (Balland et al., 2022). Relatedness, instead, measures the degree of similarity or connection between different economic activities, products, or technologies (Boschma, 2017). Balland et al. (2020) shows how both knowledge complexity and relatedness constitute key building blocks of smart specialisation policy levering regional strengths and opportunities. Additionally, there exists a growing body of literature focusing on the link between complexity and sustainability (e.g., Pugliese and Tuebke, 2019; Mealy and Teytelboym, 2022; Sbardella et al., 2022; European Commission, 2024)); and other empirical analyses have relied on the concepts of complexity and relatedness to explore the role played by digital skills on green diversification in European regions, comparing their relevance for the development of green versus non-green technologies (Santoalha et al., 2021), as well as measuring skills premium considering their degree of complementarity (Stephany and Teutloff, 2024).

In this paper we argue how incorporating complexity metrics into R&I policy evaluation processes can substantially improve the EU's capacity to design and implement effective policy interventions. Collecting consistent data on how technologies advance and innovation ecosystems evolve can provide policymakers with critical insights into regional development, helping them identify strengths and weaknesses of regional innovation ecosystems. This, in turn, can provide valuable information on how to develop a comprehensive approach to design policy interventions able not only to address short-term needs for technological progress, but also to foster long-term innovation and economic growth, thereby enhancing competitiveness against international competitors. Specifically, the objective of the paper is to provide a description of how complexity and relatedness metrics are measured and can be used to gather more strategic insights into technological development in the EU. We do so by focusing on three applications linked to different EU policy priorities: 1. Achieving technological sovereignty and strategic autonomy; 2. Deploying the European Green Deal (EGD); 3. Fostering regional R&I collaborations within the EU.

Knowledge complexity studies the geography and dynamics of innovation activities, adopting an outcome-based approach, i.e., data on regional innovation activities (such as patent data) is used to infer the presence of bundles of capabilities. Specifically, the Knowledge Complexity Index (KCI) is an indicator measuring regions/countries' innovation capacity from data connecting such regions or countries to different types of technologies present in their portfolio. Similarly, the Technology Complexity Index (TCI) measures the complexity required to patent in a given technological field.

The intuition behind these indicators is that technologies vastly differ in terms of value and growth potential. Technologies relatively easy to copy and move over space typically require a lower number of capabilities to be undertaken, thereby conferring a lower competitive advantage to the countries or regions in which they are located (Balland et al., 2020). On the contrary, more complex technologies combine a higher number of capabilities, are more concentrated in space and are characterised by a higher potential in terms of growth and overall competitiveness (Balland and Rigby, 2017). Therefore, these indicators are calculated by studying the number of countries or regions able to patent in a given technological field, and infer the quality of a country or region's knowledge base by looking both at the technology fields in which it is able to specialise and at the other places where those technologies are also present (Balland and Rigby, 2017; Hidalgo, 2021).

Close to knowledge complexity is the concept of technological relatedness. Two technologies are considered related when they rely on the same knowledge and competencies to be produced (Hidalgo et al., 2018; Balland et al., 2019). Relatedness, thus, measures the "similarity" between the know-hows underpinning different technologies. This similarity is inferred once again exploiting information on the geographical location of innovation activities: two technologies are similar if they are often produced in the same places. This information is then used to assess which technologies are feasible for that region or country to develop. Therefore, relatedness provides information on the technological potential of a country or region in a given technology, as it refers to the costs that a country or region has to sustain when moving into a new technology (Boschma, 2017; Hidalgo et al., 2018). Intuitively, the more related current and new technologies are, the lower the cost to specialise in the new field. It follows that it is relatively easier to diversify in technologies requiring capabilities that largely overlap with those already present in a country or region. On the contrary, when the overlap between existing and new capabilities is small, jumping into a new technology field becomes more risky and costly (Bachtrogler-Unger et al., 2023).

Recent geopolitical and economic developments have shifted the European policymaker's agenda towards three main priorities: Strategic autonomy, global decarbonization and economic efficiency (Aghion et al., 2023). Against this backdrop, the role of knowledge complexity in the pursuit of these three priorities will be discussed below.

Technological sovereignty and strategic autonomy: With the current geopolitical shifts and global economic instability in mind, policymakers have to balance the pursuit of economic efficiency while ensuring economic and geopolitical resilience. The position of the EU to lead technological change in areas related to key strategic technologies remains weaker than its international counterparts due to its limited existing knowhow to develop specializations within these technologies (European Commission, 2024). To further develop the necessary technological capacities and to guarantee access to critical technologies, the EU could simultaneously build upon existing "in-house" knowledge and international sources of knowledge (e.g., via collaborations, exports or FDI). Tapping into these external networks, increases the ability to learn or gain access to relevant capabilities, necessary to develop and diversify into new types of technologies and to further stimulate and sustain local knowledge creation (Boschma, 2005).

Nevertheless, betting on the right strategic technologies and connecting the right places to develop them is becoming an increasingly difficult exercise due to the speed of technological change and the lack of data to inform such decisions. With too many technologies and global knowledge ecosystems to intuitively assess in terms of optimal investments, complexity can provide directionality in R&I policymaking. More precisely, it can be used as a tool to identify the types of strategic technologies in which the EU could leverage its existing capabilities for further specialization by assessing existing technological competencies and future technological potential. Furthermore, it can help identify technological complementarities with other countries, thereby providing insights on how to manage the quality and depth of the external relations.

Deploying the European Green Deal: The latter is also relevant to effectively deploy the European Grean Deal. The EGD aims to offset greenhouse gas emissions by 2025 while enhancing economic growth. To meet carbon neutrality goals, the EU will have to accelerate the development of climate-related technologies as climate targets cannot be met by only relying on existing technologies. In this case, complexity and relatedness metrics can provide guidance regarding the direction of policy intervention by evaluating which green technologies have the potential to be developed in the EU and which areas are better placed to do so based on their existing capabilities. This type of analysis can thus provide insights into identifying territorial investment opportunities to develop a particular green technology and on which green technologies the EU should be focussing on.

The European R&I connectivity network: Technological and innovative capabilities can vary substantially across Europe. In many cases, technologies are characterised by a specific geography, but existing capabilities may be diversified and expanded through inter-regional collaboration. Nevertheless, this type of collaborations remains sparse, especially in Europe (Balland, 2022) where within-country and even within-region collaboration has a strong preference despite EU cohesion policy aiming to close the research and innovation divide between countries and regions. In this case, knowledge complexity can provide direction into discovering the untapped potential in inter-regional or cross-country collaborations in Europe, even in regions that may not come to mind as quickly. It allows for mapping the potential of European regions to develop certain technologies, which can then be used to identify technologies that best fit a specific regional ecosystem. This overview would allow policymakers to promote greater cohesion throughout Europe.

REFERENCES:

- Aghion, P., Ahuja, K., Bown, C. P., Cantner, U., Criscuolo, C., Dechezlepretre, A., Dewatripont, M., Hausmann, R., Lalanne, G., McWilliams, B. et al. (2023), Sparking europe's new industrial revolution-a policy for net zero, growth and resilience, Technical report, Bruegel.
- Bachtrogler-Unger, J., Balland, P. A., Boschma, R., Schwab, T., (2023), Technological capabilities and the twin transition in Europe: Opportunities for regional collaboration and economic cohesion, Bertelsmann Stiftung, Austrian Institute of Economic Research, Utrecht University, Artificial and Natural Intelligence Toulouse Institute, University of Stavanger.
- Balland, P. A., (2022), Innovation Policy for a Complex World, Chapter 14 in: Science, Research and Innovation Performance of the EU 2022, European Commission's Directorate-General for Research and Innovation, Publications Office of the European Union, Luxembourg.
- Balland, P. A., and Rigby, D. (2017), 'The geography of complex knowledge', Economic geography, 93(1), 1–23.Balland, P. A., Boschma, R., Crespo, J., and Rigby, D. L. (2019), 'Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification', Regional studies, 53(9), 1252-1268.

- Balland, P. A., Broekel, T., Diodato, D., Giuliani, E., Hausmann, R., O'Clery, N., Rigby, D. (2022), 'The new paradigm of economic complexity', Research Policy, 51(3).
- Balland, P. A., Jara-Figuerosa, C., Petralia, S. G., Steijn, M. P. A., Rigby, D. L., Hidalgo, C. A., (2020), 'Complex economic activities concentrate in large cities', Nature Human Behavior, 4 (3), pp. 248-254.
- Boschma, R. (2017), 'Relatedness as driver of regional diversification: A research agenda', Regional Studies, 51(3), 351-364.
- Boschma, R., (2005), 'Proximity and innovation: a critical assessment', Regional Studies, 39(1), pp. 61-74.
- Cavicchi, B., Peiffer-Smadja, O., Ravet, J., Hobza, A., (2023), The transformative nature of the European framework programme for research and innovation: analysis of its evolution between 2002-2023, European Commission, Directorate-General for Research and Innovation, Publications Office of the European Union.
- European Commission (2024), Science Research and Innovation Performance Report - forthcoming
- Hidalgo, C. A. (2021), 'Economic complexity theory and applications', Nature Reviews Physics, 3(2), 92–113.
- Hidalgo, C. A., Balland, P.-A., Boschma, R., Delgado, M., Feldman, M., Frenken, K., Glaeser, E., He, C., Kogler, D. F., Morrison, A. et al. (2018), 'The principle of relatedness', in Unifying Themes in Complex Systems IX: Proceedings of the Ninth International Conference on Complex Systems, Springer, pp. 451–457.
- Hidalgo, C. A., Hausmann, R., (2009), 'The building blocks of economic complexity', a, 106 (12), pp. 10570-10575.
- Mealy, P., Teytelboym, A., (2022), 'Economic complexity and the green economy', Research Policy, 51(8).
- Pintar, N., Scherngell, T., (2021), 'The complex nature of regional knowledge production: Evidence on European regions', Research Policy, 51(8).
- Pugliese, E., Tuebke, A., (2019), Economic complexity to address current challenges in innovation systems: a novel empirical strategy for regional development Industrial R&I – JRC Policy Insights.
- Romer, PM., (1990), 'Endogenous Technological Change', Journal of Political Economy, 98(5), pp. 71-102.
- Santoalha, A., Consoli, D., Castellacci, F. (2021), 'Digital skills, relatedness and green diversification: A study of European regions', Research Policy, 50(9).

- Sbardella, A., Barbieri, N., Consoli, D., Napolitano, L., Perruchas, F. Pugliese, E., (2022), The regional green potential of the European innovation system (No. JRC124696). Joint Research Centre.
- Stephany, F., Teutloff, O., (2024), 'What is the price of a skill? The value of complementarity', Research Policy, 53(1).
- Storper, M., Venables, A., (2004), 'Buzz: face-to-face contact and the urban economy', Journal of Economic Geography, 4(1), pp. 351-370.

Mart Laatsit, Magnus Nilsson, Asa Lindholm Dahlstrand
Lund University, CIRCLE, Sweden

CAPTURING THE SYSTEM-LEVEL EFFECTS OF INNOVATION POLICY

ABSTRACT

Two trends illustrate the current state of affairs in innovation policy studies. On one hand, the changing rationale of innovation policy towards addressing grand societal challenges (Schot and Steinmueller, 2018; Weber and Rohracher, 2012) has led to calls for more systemic innovation policy instruments (Janssen, 2019; Smits and Kuhlmann, 2004; Wieczorek and Hekkert, 2012). On the other hand, the rise of these systemic instruments for transformative change brings new challenges for policy analysis and evaluation (Grillitsch et al., 2019; Haddad and Bergek, 2020), calling for novel approaches for assessing the role and performance of individual instruments in the broader policy-mix. In particular, previous research has shown that an assessment of a policy programme should consider effects on different levels and of different types (Borrás and Laatsit, 2019; Dosso et al., 2018; Gök and Edler, 2012). Therefore, we need more comprehensive ways for assessing the merits of individual innovation policy instruments from a system perspective.

We argue that this requires an analytical framework that covers the three types of effects embedded in policy instruments: first-order, second-order and system-level effects. Firstorder effects are firm-level effects directly on the targeted firms while second-order effects are firm-level effects on third parties not targeted by the policy activity. System-level effects, however, are an effect on the system itself and its functions. Both first-order and secondorder effects have been extensively studied and conceptualised (Hottenrott et al., 2017). However, the increasing interest in mission-oriented and transformative innovation policies has brought a renewed attention to studying systemic change induced by policy (Amanatidou et al., 2014; Arnold et al., 2018; Haddad et al., 2022; Janssen, 2019; Kao et al., 2019; van Mierlo et al., 2010; Warwick and Nolan, 2014).

Earlier studies have discussed assessing system-level effects, but mostly as produced by the whole policy-mix. Examples of this include 'systems of evaluations' (Arnold, 2004; Jordan et al., 2008) as well as 'meta-analyses' (Edler et al., 2008; Magro and Wilson, 2013), both combining different analyses for a systemic understanding. We argue that it is not always possible, nor necessary, to take a policy-mix perspective to analyse the system-level effects of innovation policy, as it can also be done on programme level. Therefore, we take a programme-centric view and demonstrate how system-level effects can be identified for an individual programme.

We propose a conceptual framework for capturing the first-order, second-order and systemlevel effects of a single policy instrument. We assess the approach through an analysis of the Swedish Innovation Agency VINNOVA's Innovative SME programme. Our

analysis includes a combination of quantitative and qualitative studies. First, we collected information and data for the population of 1341 SMEs supported through the programme between 2001 and 2015. A further cross-sectional study was conducted on a sample of 60 companies. The results demonstrate a clear link between VINNOVA support and the system's functions and functionality. We also found a solid relationship between the financial support a company received and the width of the number of affected functions. The findings strongly support the assumption that the funding contributes to the companies' ability to both develop and influence system functions.

The paper makes three important contributions. First, it demonstrates on a conceptual level how the three types of effects (first-order, second-order and system-level) can be combined to evaluate a policy programme. Second, it develops an analytical framework for assessing the system-level effects of a single programme, based on the functions of innovation systems. Third, it assesses the framework empirically using a novel dataset on Swedish SME support programmes.

We proceed as follows: first, we introduce the theoretical framework. Second, we provide an overview of the data and methodology. Third, we present the results and three case studies of Swedish companies. We conclude with a discussion on policy implications and further research perspectives.

Mart Laatsit, Markus Grillitsch , Harald Rohracher
Lund University, CIRCLE, Sweden
Lund University, CIRCLE, Sweden
Linköping University , Sweden

ASSESSING THE TRANSFORMATIVE POTENTIAL OF POLICY INSTRUMENTS: THE CASE OF SWEDEN

ABSTRACT

Transformative innovation policy (TIP) implies increased attention to the entire mix of policies required to stimulate one or more focal socio-technical system(s) to develop in a desired direction. When designing such mixes, it is important to understand to what extent different innovation policy instruments – established as well as emerging – have the potential to contribute to system transformation. Researchers have suggested several ways to assess to what extent individual instruments or entire policy mixes are (or have the potential to be) transformative, based on the main characteristics of TIP as described in the literature (Borrás and Schwaag Serger, 2022; Laatsit et al., 2022; Salas Gironés et al., 2020), but so far the practical experience with these frameworks is highly limited. This restricts the possibility to develop comprehensive and comparative empirical knowledge about how different instruments can be combined into a coherent and consistent policy mix for system transformation.

The study builds on the conceptual approach first outlined by the authors in Laatsit et al. (2022). This approach studies the transformative potential of innovation policy instruments combining the characteristics of innovation policy instruments with an assessment of their transformative potential. The latter is based on the four transformative failures, first suggested by Weber and Rohracher (2012), which has become a common reference point in analyzing transformative processes and policies.

We seek to pilot the conceptual model in the context of Swedish transformative innovation policy. We explore its use for assessing the transformative potential of the Swedish innovation policy mix and the extent to which this potential has been used through policy design and implementation. As an example, this would include answering to what extent and how directionality and reflexivity are included in the current instruments, and whether the instruments allow for sufficient coordination across sections and levels of government.

The paper further builds on the understanding that current innovation policy practices are often characterised by layering (Kivimaa and Kern, 2016) where a new framing of policies is layered on top of older practices. This can lead to situations where the transformative agenda is dominating the policy discourse, but the actual evaluation practices are still based on older innovation policy paradigms (Rohracher et al., 2023).

Recent conceptual work by Laatsit, Grillitsch and Fünfschilling (2022) has shown that the traditional innovation policy instruments hold a significant potential for addressing the transformative failures of directionality and demand articulation. The potential to address the coordination and reflexivity failure is smaller, but can be compensated for by making use of novel (or reinvented) instruments. These include mission-oriented innovation policies, instruments encouraging experimentation, regulatory policies, as well as new forms of evaluation and governance.

The study is implemented in the following steps:

- First, we develop the research design interactively with policymakers at Vinnova to finetune the research focus and select target instruments to include in the analysis.
- Second, we gather data on policy instruments and mixes through interviews or focus group meetings with Vinnova and other policymakers as well as desk-research using publicly available data on policy instruments. The data will be analyzed through qualitative content analysis.
- Third, we arrange a seminar/workshop with Swedish policy makers where we present and discuss the results of the analysis, with the aim to better understand the challenges of realizing the transformative potential of innovation policy instruments and mixes. We will also discuss how innovation policy can be designed and implemented in a way that enables the realization of its transformative potential.

Michael Dinges, Klaus Schuch
AIT Austrian Institute of Technology, Austria

DESIGNING TAILORED IMPACT PATHWAYS FOR MISSION-ORIENTED R&I POLICIES

ABSTRACT

The implementation of national and regional initiatives to support the Horizon Europe missions is of paramount importance for achieving the ambitious goals set by Horizon Europe. Therefore, Austria has established Mission Action Groups (MAGs) responsible for developing mission-specific action plans at a national level. These groups, initiated by the Austrian Federal Ministry of Education, Science and Research (BMBWF), are supported by the Mission Facility for Policy Learning, Foresight Monitoring, and Evaluation. This facility aids in the strategic planning and implementation of these missions by developing Impact Pathways that facilitate continuous monitoring and evaluation of progress in the course of the implementation of the EU Missions in Austria.

The process of developing impact pathways is currently ongoing and involves several steps, including:

- Literature Review: A comprehensive analysis of potential Impact Pathways based on strategic EU mission documents, MAG implementation plans in Austria, and scientific literature.
- Defining goals and Stakeholders: The initial phase involves defining the goals, target groups, and purposes for elaborating impact pathways in consultation with the EU mission leaders and MAGs. This includes identifying key actors and data sources necessary for creating efficient and effective Impact Pathways
- Stakeholder Engagement: Exploratory interviews and focus groups with stakeholders to gather insights and perspectives on the notion of Impact Pathways. The synthesis of these interviews helps identify key aspects and common requirements for the pathways.
- Impact Pathway Design and Refinement: Defining qualitative success criteria and indicators for each EU mission, aligning them with specific mission goals and relevant areas. Workshops are held to discuss and refine the draft pathways based on feedback.
- Gap Analysis: Identifying gaps within and between the implementation plans and impact Pathways, followed by strategic discussions to address these gaps in workshops.

Within a Program Theory, an Impact Pathway describes the expected causal linkages between an intervention and its expected outcomes. As Douthwaite et al. (2003) describe, "A program theory consists of a sequenced hierarchy of outcomes—in other words, an

impact pathway with milestones on the route. The hierarchy begins with the project outputs, followed by a chain of intermediate outcomes that are then followed by the wider and often longer-term outcomes." Alvarez et al. (2010) emphasize the importance of identifying behavioral changes among targeted stakeholders (e.g., end-users, policymakers) as a result of the intervention. This highlights the critical role of stakeholder engagement and the anticipated changes in their behavior due to the implemented actions. Muhonen et al. (2020) focus on the "productive interaction" between science and society, allowing societal actors to influence scientific actors, thereby creating new kinds of scientific value, and vice-versa. This interaction is a dominant mechanism through which impact is achieved, emphasizing the co-creation and collaborative aspects of the Impact Pathways. Reed et al. (2021) view Impact Pathways not as a chronological sequence but as a productive configuration of actors and contexts where outputs, intermediate steps, and outcomes are achieved. This approach frames the pathways within five key impact areas where the intensity of impact can be assessed. Belcher et al. (2020) define Impact Pathways based on primary actors or the actions to be influenced: For example, a policy pathway illustrates the constellation of actors and measures theoretically leading to a policy change.

Overall, the depiction of Impact Pathways aims to illustrate how planned activities trigger a sequence of events leading to the desired outcomes. Ultimately, the pathways provide a set of assumptions or hypotheses necessary for achieving the goals, which can be tested through monitoring and evaluation activities. Visualizing the impact development through various, potentially interacting pathways enables stakeholders to anticipate how inputs and activities within their control relate to immediate outputs and, subsequently, to the outcomes and broader impacts at the system level (Wittmann et al. 2022). Within the schema of the Impact Pathways, only inputs and activities, and outputs can be controlled. While outputs and outcomes are within the influence of the missions, their interest extends beyond this scope.

As a concept for monitoring and evaluation (see Bruno and Kadunc 2019), but also for increasing impact orientation of R&I policies, the notion of impact pathways has gained considerable attention in Horizon Europe. The Horizon Europe's impact pathways are designed to ensure that the funding leads to tangible benefits for society, the economy, and science. These pathways are typically categorized into three main types: scientific impact, societal impact, and economic impact. Among the 9 Horizon Europe impact pathways, delivering benefits and impact through EU Missions is one storyline, for which pathways to impact have only been outlined at a very aggregate level, if at all. The novelty of our approach lies in the development of specific Impact Pathways tailored to the EU Missions and its implementation at a national level, while ensuring some degree of comparability between them. This is achieved through a structured methodology that integrates mission-specific goals and activities while elaborating a common framework for monitoring and evaluation. The Impact Pathways are designed to:

- Show the implementation pathways towards mission goals, i.e. clearly outline how measures and activities lead to desired results.
- Enable comparability: i.e. develop prototype pathways applicable to all EU missions at a national level, adaptable to specific needs.
- Provide a foundation for monitoring and evaluation: Create a basis that aligns with the European Commission's monitoring and evaluation framework.

Impact Pathways for mission-oriented R&I policies can be clustered along specific impact domains: 1) 1) Science and Technology, 2) Economy and Society, 3) Policy and Governance, and 4) Environment and Health. Each domain has prototypical pathways to be followed, from which tailored pathways can be elaborated together with the stakeholders engaged in programming and monitoring the implementation of mission-oriented policies. The focus of the paper is on how these tailored Impact Pathways can be designed, emphasizing stakeholder engagement and iterative refinement.

The systematic development and implementation of Impact Pathways for the EU missions in Austria represents a strategic approach to addressing the implementation and monitoring and evaluation of societal challenges through coordinated efforts in science, technology, policy, and governance. The development of Impact Pathways is expected to result in: 1) Enhanced Clarity and Transparency: Clearly telling a story about defined linkages between goals, actions, and outcomes, making the process of achieving objectives more transparent and understandable. 2) Improved Learning and Adaptation: Continuous evaluation and adaptation of Impact Pathways to learn from experiences and enhance the effectiveness of measures. 3) Increased Efficiency and Effectiveness: A systematic analysis of Impact Pathways enables targeted resource allocation, identification of gaps, and development of measures to improve goal attainment. One of the primary challenges to be tackled in developing Impact Pathways is ensuring coherence and integration across diverse mission objectives, type of actions pursued, and sectors addressed. This requires effective coordination among various stakeholders, clear communication, and a flexible approach to incorporate new insights and feedback. The development of impact pathway prototypes and a strong stakeholder engagement process is expected to provide the means for iterative reflection and refinement that contributes to the advancement of the implementation of the EU Missions in Austria on the one hand, while paving the way for continuous monitoring and evaluation on the other. The impact pathways for the implementation of the EU missions in Austria will be drawn up based on the developed concept between June and September 2024. Therefore, we can actually test the concept empirically until the ReValuation'24 conference using selected EU missions (e.g. 'EU Mission Water', 'EU Mission Soil') and present and discuss the results "fresh from the field".

REFERENCES:

- Alvarez, S., Douthwaite, B., Thiele, G., Mackay, R., CÛrdoba, D., & Tehelen, K. (2010). Participatory impact pathways analysis: a practical method for project planning and evaluation. Development in Practice, 20(8), 946-958.
- Douthwaite, B., Alvarez, S., Cook, S., Davies, R., George, P., Howell, J., ... & Rubiano, J. (2007). Participatory impact pathways analysis: a practical application of program theory in research-fordevelopment. Canadian Journal of Program Evaluation, 22(2), 127-159.
- Bruno, N., & Kadunc, M. (2019). Impact Pathways: Tracking and communicating the impact of the European Framework Programme for research and innovation. fteval Journal for Research and Technology Policy Evaluation, (47), 62-71.
- Belcher, B. M., Davel, R., & Claus, R. (2020). A refined method for theory-based evaluation of the societal impacts of research. MethodsX, 7, 100788.
- Kroll, H., Hansmeier, H., & Hufnagl, M. (2022). Productive interactions in basic research an enquiry into impact pathways at the DESY synchrotron. Technological Forecasting and Social Change, 175, 121408.
- Wittmann, F., Hufnagl, M., Roth, F., Lindner, R., & Kroll, H. (2021). Ein Rahmen f¸r die formative Evaluation und Wirkungsmessung von missionsorientierten Innovationspolitiken. Abschlussbericht der wissenschaftlichen Begleitforschung zur deutschen Hightech-Strategie, 2.

Katharina Warta, Michael Stampfer
Technopolis, Austria
Vienna Science and Technology Fund (WWTF), Austria

A SCIENCE FUNDING ORGANISATION, A SEVERE SOCIETAL CRISIS, AND A VERY RAPID RESPONSE

ABSTRACT

Organizations (RFOs) adapted by offering specific grants within faster procedural frameworks. However, as this is a rule-based business, many RFOs needed some time to change track. The following example from Austria shows that track change can be achieved also within a very short period of time: Two weeks in early Spring 2020 for a complete cycle: Specific design of the funding initiative, issuing of the call, writing of the applications, composition of a funding jury, selection, funding decision, start of the projects. Can we see impacts from such a quick shot of a comparatively small RFO, and if yes, do we find them also in changed practices in the field, or even adaptations in positioning of the RFO itself?

The Vienna Science and Technology Fund (WWTF) has commissioned Technopolis Austria in 2023 to evaluate this funding initiative, the COVID-19 Rapid Response Call. The evaluation looks at the process, implementation, results, and impact of the Call and formulates formative conclusions and recommendations, in particular on the extent to which lessons can be drawn from the experience with the COVID-19 Rapid Response Call for other programmes and processes of the WWTF or other RFOs. The evaluation is based on document analysis, in particular coded project final reports, a selfassessment report by the WWTF, a network analysis of collaborations, a focus group with representatives of funded projects, as well as interviews with applicants, jury members, and various stakeholders. We also provide an overview of international experience with comparable programmes.

Overall, the collected evidence reveals a very positive picture of the WWTF's COVID-19 Rapid Response Call, with an extraordinarily high scientific and societal impact in terms of financial expenditure and speed. The call had a mobilizing effect on the Viennese research community because the WWTF suddenly and quickly opened up scope for action and allowed the real-time collection of data by social and health scientists. The central prerequisite for this is the specific constellation of the WWTF – small, independent, a fund, trust capital accumulated over many years, good networking, and strategic expertise.

A special feature of the Call was the speed with which the WWTF designed and implemented it, which was also made possible by the involvement of organisational leaders such as university rectors in the pre-selection of proposals. The speed of the design

and selection process was key to the success of the funded projects, with contributions to societal impact and science being made more quickly and to a greater extent. The objective of enabling the rapid collection of data for research purposes was thus achieved. The WWTF has also taken the experience gained from the Call into account, for example by making the format of the "Additional Funding Measures" even more flexible in 2021 and by increasing the maximum funding amount, which has since already been used for three other initiatives. In addition, the WWTF subsequently launched two new major calls, in which the experience from the COVID-19 Rapid Response Call could be incorporated ("Empirical Social Sciences") or provided an impetus for it ("Public Health"). 24 projects have been awarded each up to € 50.000 for data collection and some experimental work. The direct results of the COVID-19 Rapid Response Call were 83 publications directly related to the funding, of which 53 were peer-reviewed and 54 open access publications. In addition, 19 indirect publications were recorded, of which 15 were peer-reviewed. Furthermore, 40 new academic collaborations were established, eleven international and three national ones. 23 projects reported at least one successful follow-up application. Three scientific career steps were recorded.

Many funded projects were able to make unusually high and rapid impact contributions, particularly in the area of pandemic management and control. Examples include contributions to the development and commercialization of SARS-COV-2 antibody tests, the successful launch of the "Alles Gurgelt" (PCR mass tests) project in Vienna, where upscaling was supported, and the development and implementation of the Corona traffic light system, which was supported by several projects. WWTFfunding has also contributed to providing decision-makers with information on the course of the pandemic and the impact of measures taken on those affected. From a scientific point of view, the projects had an advantage in the competition for scientific publications and citations because the results were available very early and the content of the first two waves of the coronavirus pandemic had already been taken into account. A major difference from other research projects was the high level of media interest in the research projects and the programme, which was generated comparatively quickly after the start of the project.

For many of the researchers involved, this funding experience was characterized by the impression of being able to make an unusually large contribution to solving a very specific societal problem. This had a strong mobilizing effect, which was also reflected in the target groups of the panel surveys. This was also associated with new collaboration patterns and a markedly increased workload. Work that would normally take months was compressed into a few days or weeks. This happened at a time when the actual research work was sometimes made more difficult by the working conditions during the first lockdown

In our interactive presentation we will focus on two findings of this evaluation and the call itself:

1. Rapid response is possible but for some it might be easier than for others
2. Relevance for missions / challenges is possible for a funder of scientific research, but certain framework conditions and pathways are necessary requirements

Ad 1: For many years, the WWTF has been highly recognized and trusted by researchers and stakeholders for attracting and selecting high-quality research proposals, in time-consuming procedures, and subsequently supporting their implementation. Against this background of extensive evaluation experience, an exception could be made to the procedure in the event of a crisis, to select projects quickly and still create sufficient legitimacy. The main added value of the WWTF's COVID-19 Rapid Response Call was indeed that the funded researchers were able to quickly carry out research on relevant topics through a project on an institutionalized ground. An earlier start – April 2020, in the first wave – increased the impact both at the scientific level (publication success) and at the political and societal level (visibility and early provision of action-oriented data). A number of WWTF's features have also helped here: small size, local embeddedness, nature as a private non-profit funding organization, flexible structures.

Ad 2: With the COVID-19 Rapid Response Call at the latest, we argue that the WWTF collected valuable experiences regarding "third generation research governance" (Arnold and Barker, 20225), which is oriented towards the impact of research on major societal challenges. One of the key success factors here is that the WWTF has clearly defined concrete objectives that are reflected in the selection criteria and also facilitate the evaluation of impact and achievement of objectives. For WWTF the issue of societal relevance has become more important over the last years. Various of its current initiatives do not stop with opening up calls for researchers and waiting for them to apply but actively try to co-shape communities and practices.

The presentation at the REvaluation conference will be structured along these two elements – speed and relevance – and from two perspectives: Katharina Warta, responsible for the evaluation, will present evidence collected and analysed with the evaluation, Michael Stampfer, CEO of the WWTF, will shortly give an overview how the experience fed into new calls. He will also examine the question of whether and how a science fund can become involved in urgent situations, and in which situations - despite societal urgency - it should refrain from doing so, not least to ensure independence and quality.

The experience with the COVID-19 rapid response call showed that exceptional speed is possible in exceptional circumstances, without any accidents and with lots of outcomes and impacts. The topic “data” provided the necessary focus and a tangible objective to stay on track. Further, funders of scientific research can contribute meaningfully to societal challenges, as long as the objectives are clear and everybody knows their roles. Given the exceptional situation in March 2020, the WWTF certainly benefited from accumulated trust in the procedures. It seems that the trust capital invested has been returned manyfold. Also, on the research organisations and universities’ side, a high level of trust and engagement was perceived. The specific focus on data helped as it played the role of a facilitator to boost preexisting local data practices.

Wolfgang Polt, Michela Magas, Louise Staffas, Anders Holmgren
Joanneum Research POLICIES, Austria
Industry Commons Foundation, Sweden
Formas Research Funding Agency, Sweden
Vinnova, Sweden's Innovation Agency, Sweden

EX-ANTE ASSESSMENT OF MISSION-ORIENTED PROGRAMMES: THE CASE OF THE SWEDISH IMPACT INNOVATION PROGRAMME

ABSTRACT

In a number of countries as well as at the EU level, recent years have seen attempts to design and implement 'transformative mission-oriented policies' (see for an overview Edler, Matt, Polt & Weber 2024 forthcoming), i.e. policies and programmes which aim at systemic changes and hence are transgressing traditional borders of administrations, funding, actor constellations and sectors. Setting up such programmes has proven to be difficult and is – even in the most advanced countries - still in experimental stages and is very much 'path-dependent' on existing 'policy trajectories' of the respective countries (Larrue 2021, Polt & Weber 2023). Against this background, it is very interesting to study examples of such experiments – even more so if they are on a considerable scale, as is the case of the Swedish 'Impact Innovation' programme, which aims to "take on the greatest societal challenges of our time and of the future" trying to bring together "all corners of society to be able to make a big difference".

The Impact Innovation (II) programme has been launched against the following background and in the following context: Sweden has in the past couple of decades seen a continuous development of successive generations of thematic, challenge-driven and mission-oriented innovation programmes. The previous generation (the Strategic Innovation Programmes or SIPs) already had a thematic orientation, but were not able to catalyze transformation in the way and extent which had become apparently necessary during the time of the programs. From that observation, came the decision to design a mission-oriented programme. Based on intense internal reflections on how to implement mission-oriented polices (see e.g. Hill et al 2022) and also on the findings of previous evaluations of the SIPs (see for a meta-evaluation Åström, T. & Arnold, E., 2023), Impact Innovation was conceptualised.

Mission-oriented policies are implemented in their respective regional or national contexts. While they share some characteristics (e.g. the increasing necessity for whole-of-government and whole-of-society approaches), the governance structures, actor constellations and mix of instruments differ considerably. In Sweden, as compared to other countries, there is a greater role of agencies, the collaboration of which is the main leverage for cross-sectoral government action. E.g. for the II programme, three of the most

important funding agencies in Sweden, Formas, the Swedish Energy Agency and Vinnova, teamed up. While the role of central political level remained limited, the role of regions as actors in implementing missions is considerably larger in Sweden than in other comparable countries and hence the representation of regions in the respective consortia was very high. This may lead to novel solutions for regional innovation as reflected in the upcoming Regional Innovation Handbook, developed in collaboration with the European Committee of the Regions and the New European Bauhaus initiative digiNEB, which takes lessons from the Swedish Vinnova Duved project and the Norrlands Model developed as part of the Vinnova, Formas and Swedish Energy Agency project "Visions in the North". In our contribution, we will describe the influences of the context to the programme development and design in greater detail with a view to comparisons to other countries (see final paragraph).

The II programme had a very significant bottom-up element: consortia of a broad range of stakeholders (large and small businesses, public sector organizations, research institutions and civil society) were asked to formulate joint visions in relation to three designated challenge areas and could decide on their own how to structure their respective mission programme. They were supported also in the pre-proposal stage to form their consortia (which was a labour intensive effort). They were asked to propose governance structures for their collaboration in the form of programme offices which would also be funded through 'Impact Innovation' and should meet certain criteria with respect to composition, administrative capacity etc. These methods were developed in the context of Hill, et al. (2022), in which the authors advocate for innovative approaches such as the 'snowball effect' methodology, and the impact this may have on the implementation of innovation programmes.

There was a very intense selection process, involving experts from different strands (academia, business, civil society) reflecting also the breadth of actors in the consortia. Evaluators were informed and trained to be able to handle the transformative and mission-oriented nature of II which asked for different approaches to assessment – especially going beyond the traditional 'excellence' criteria and including the 'transformative potential' of the respective proposal. Several rounds of assessment took place leading to joint recommendation from the evaluators. The final decision was taken from a portfolio perspective by the funding agencies, reflecting among other things, that the program portfolio as a whole should cover the three challenge areas' representation and sectoral and actors' representations. Portfolio selection was, however, only applied on proposals with enough quality, as assessed by the reviewers. We will describe the

experiences and challenges of this selection process from a hands-on perspective.
Underlining the experimental character of the approach, and the difficulties to come up with predefined metrics applicable across the board, it was left to the consortia to define goals and related KPIs and milestones. For some consortia, the concrete formulation was even left to the initial phase of the project. Hence, also monitoring and evaluation of the progress and interim and final outcomes of the different missions will require very knowledgeable evaluators and intense support from the funding agencies. It is a shared perception among the funding agencies that the novelty of the approach also requires novel ways to work with the consortia – very different ones from traditional 'fire and forget' funding decisions in research funding projects and more in the form of frequent formative follow-ups as well as formal evaluations.
Finally, we would compare the approach taken in Impact Innovation to other countries, especially Austria, with its governance structures for implementing the EU missions, with Germany and the missions in its High-tech strategy, the Netherlands and the missions-oriented programmes in its top-sector strategy, and the programmes in Norway run e.g. Pilot-e ….) and try to draw lessons about the potentials and the challenges of the respective approaches.

REFERENCES:

- Åström, T. & Arnold, E., 2023. Meta-evaluation of the Swedish Strategic Innovation Programmes As policy instruments for industrial competitiveness and systems innovation. VR 2023: 16 ed. Stockholm: Vinnova.
- Edler, J., Matt, M., Polt W. & Weber M. 2024 Transformative mission-oriented STI policy: Theoretical and conceptual rationales, intervention logics and challenges of an emerging type of STI. In J. Edler, M. Matt, W. Polt, & K. M. Weber (Eds.), Transformative Mission-Oriented Innovation Policies. Edward Elgar Publishing forthcoming
- Hill, D. et al. 2022. Designing Missions. Mission-oriented innovation in Sweden - A practice guide from VINNOVA.
- Larrue, P. (2024). Challenges and opportunities of mission-oriented innovation policies for realizing transformative agendas: General principles and application to the case of the fight against climate change. In J. Edler, M. Matt, W. Polt, & K. M. Weber (Eds.), Transformative Mission-Oriented Innovation Policies. Edward Elgar Publishing (forthcoming)
- Larrue, P., 2021. The Design and Implementation of Mission-Oriented Innovation Policies: A new systemic approach to address societal chalenges, OECD Science, Technology and Industry Papers No 100. Paris: OECD.

- Nylander, J. Polt, W. & Weber, M., 2024. Mission Playbook - Stories from Practice. TRAMI / VINNOVA [https://www.trami5missions.eu/sites/default/files/2024-04/TRAMI%20Missions%20Playbook_final_non-accessibiliy-ver-compressed.pdf]
- Polt, W. & Weber, M., 2023 First observations on the current state of mission implementation. TRAMI Policy Brief No. 1 https://www.trami5missions.eu/sites/default/files/2024-03/TRAMI%20Policy%20Brief%20No%201%20230228.pdf

Soohong Eum
Centre for Science and Technology Studies, Leiden University , Netherlands

RESEARCHERS' RESPONSES AND STRATEGIES TO CHANGES IN MISSION-DRIVEN FUNDING SCHEMES: A CASE STUDY OF SOUTH KOREA'S GREEN MISSION

ABSTRACT

To achieve policy objectives, such as solving societal challenges, the government sector is becoming active in intervening to stimulate relevant research and innovation activities and steer them in desired directions (Braun 2003; Whitley 2007, 2014). Recently emerging global-level challenges like climate change and pandemics, now focal points on the policy agenda, demand a greater contribution of science and technology to tackle them effectively and remain relevant to society. These trends necessitate coordinated cross-sectoral efforts that integrate various fields and prompt policy initiatives shaped by mission-oriented approaches tailored to address the specific challenge at hand (Larrue 2021).

Given that the mission requires collaboration across the entire research and innovation value chain, its associated science, technology, and innovation policies entail investments not only in applied research and development focused on economic and industrial outcomes but also in basic and fundamental research for knowledge production. Funding stands out as a recognized means to enact these endeavours. The government aims to steer the direction of research and innovation by delineating strategic priorities and setting specific targets within funding schemes, which encompass selected aims, goals, expected outcomes, and impacts. Efforts to implement this effectively have led to the establishment of thematically oriented R&D programs and the bolstering of R&D intensity through the targeted allocation of public resources to them. The diverse procedures inherent in these programs, including resource allocation and outcome evaluation, exert influence on the knowledge production process, aligning both the conduct and content of research with societal needs.

Unintended consequences of South Korea's green mission: Through the case of South Korea's Green Growth, this study investigates the impact of the changed policy landscape driven by the mission on funding dynamics, researchers' strategies to secure funding and sustain research activities, and its consequences. Green Growth, launched during the Lee Myungbak administration (2008-2013), aimed to address climate change while fostering economic development (Presidential Committee on Green Growth 2009). Its primary objective was the cultivation of green industry sectors, with the promotion of green technologies for industrialization set as the desired policy direction. Additionally, the

government identified '27 core green technologies' that need to be strategically invested and intended to concentrate funding in these areas (Korean Government Ministries Concerned 2009). Consequently, funding allocated for green technology experienced rapid growth, accompanied by an uptick in the number of research outcomes receiving financial support. However, prior research has identified the following unintended consequences: among funded research publications, the proportion lacking clear topical relevance to the mission was notably high (Eum 2022). As funding increased during the mission period, the number of researchers publishing papers through mission-related projects rapidly grew.

However, the majority of them ceased publishing papers in the field after the mission ended (Eum 2023). This suggests that the mission could not achieve its original objectives for the following reasons: Firstly, the mismatch between funding and research outcomes meant that, despite a quantitative increase in publications, the relevant knowledge expected by the government was not sufficiently generated through funding. Secondly, policy measures aimed at attracting researchers resulted in a short-term boost in research activity but failed to foster a long-term commitment, as researchers did not pivot their research focus accordingly.

To explore the misalignment between funding and research, this study will examine funderresearcher interactions from two perspectives. The first is the configuration of funding programs and instruments for influence from funders to researchers. Funders narrow down the goals and priorities of the mission and construct their portfolios in their own way, launching new programs for prioritized areas and redistributing resources through existing programs realigned to the mission. This requires mission-fitting selection, monitoring, and evaluation mechanisms in new programs, as well as changes in rebranded existing programs. It is particularly essential to mix and orchestrate different types of programs, such as thematically oriented programs and investigator-driven grant schemes, under the mission umbrella. Therefore, it is necessary to examine whether the instruments of the programs for the green mission were appropriate.

Another aspect is the response from researchers to funders in adapting to the mission-changing environment. In response to changes in the research and funding environment, researchers use strategies to adapt and secure resources to maintain autonomy and authority over their research and to sustain their research activities (Laudel 2023). Korea's green mission, a state-led agenda, reshaped the funding environment in a top-down manner, and to secure resources from rapidly increasing green research funding, researchers had to persuade funders that their research was relevant to the mission or change their research trajectory to make it relevant. In this respect, this study will examine how the mission was translated into researchers' research agendas and what impact it had on actual research practices.

To explore and bridge the aforementioned issues, this study adopts the following three dimensions as a conceptual framework. The first dimension is the principal-agent problem (van der Meulen 1998; Shove 2003). Funders (principals) and researchers (agents) have different interests: funders expect research results that can contribute to policy, while researchers hope to continue research on topics of interest to them. Due to information asymmetry—where funders may not have sufficient knowledge about the researchers' actual research content—funders might not select the best candidates for funding, or researchers might conduct research that funders do not expect. To mitigate this problem, tools have been introduced to appropriately evaluate researchers during the selection and assessment process. However, whether these tools are adequate and sufficient to handle the rapidly increasing green research funding needs to be reviewed.

The second dimension is the mix of funding in research activities. The trend of decreasing internal and institutional funding and increasing the proportion of external, competitive, and project-based funding requires researchers to secure resources from a variety of dispersed sources rather than relying on a single type of funding. Multiple projects with different priorities and expectations from different funders are aggregated to form researchers' research funding portfolios. Typically, instead of focusing on a single project, researchers and their collaborators work on mixed portfolios where resources are co-used or redistributed within the consortium. Consequently, outcomes that blend the goals and contents of each project are produced. This study will examine how these configurations and amalgamations of funding were implemented (Aagaard et al. 2021).

Finally, this study will consider how researchers articulated their relevance to external demands in response to a changing research and funding environment. To secure resources and convince funders that their research aligns with the mission's direction, researchers construct “doable” problems that align with the social world, laboratory, and experiment. In the process, researchers may abandon unfundable research trails, initiate fundable ones, or change the direction of existing ones. In this respect, this study will examine how researchers adapted to environmental changes resulting from the green mission in terms of funding applications, research practices, and reporting of results.

Data and methodology: This study examines research funding associated with the green mission and its corresponding paper outputs at the project level. This approach is chosen because the mission necessitates targeted and coordinated action across various policy domains, resulting in a portfolio of programs rather than a singular initiative. Consequently, funding is dispersed across different programs within various governmental organizations. Furthermore, project-based funding provides a clearer delineation of goals, timeframes, and expected outcomes compared to block grants, facilitating the identification of funded projects and their research outputs.
The study used project records provided by NTIS (National Science and Technology Information), a database managing all R&D investments of the Korean government, to gather data on research funding. Search strings were constructed based on unique terms and their combinations within each technology field to retrieve projects. Project records include details such as title, abstract, year, keywords, funder, PI's affiliated research institution, budget, etc. Additionally, detailed bibliographic information on papers reported as project outcomes was obtained from the Web of Science database. This data served as a reference to assess the relevance of funded papers to green technologies.
Based on the results derived from funding and publication data, interviews were conducted with researchers and officials from universities, government-funded research institutes, and funding agencies. The questions focused firstly on the influence of research policies and resource allocation under the mission on the requirements and incentives for their research. Secondly, participants were asked about their strategies for seeking, acquiring and managing research funding, as well as reporting and evaluating outcomes in response to the changed environment. These interviews provided a qualitative dimension to the study, offering the following insights: Firstly, they enable a deeper understanding of the nuanced ways in which research policies and resource allocation under the green mission influence the strategies of researchers. Secondly, the interviews shed light on the practical challenges and opportunities encountered by researchers and officials in seeking and managing research funding within the context of the mission.

While the increase in applied research and academy-industry links resulting from government intervention to achieve policy goals has been widely discussed in the literature, there has been relatively little empirical investigation into its epistemic effects, including the impact on publications, which are one of the common forms of knowledge produced. This study will tentatively present the following results: First, compared to the rapid increase in funding according to the mission, the capacity of funders to manage it has not sufficiently increased. Second, the existing program processes were not sufficiently aligned to ensure the relevance of the funded research to the mission. Third, researchers adopted the strategy of persuading funders that their research was relevant to the mission, even if it was not, by considering the timeframe of the mission and the capacity of funders. These results suggest that the dynamics corresponding to the external environment of epistemic communities should be considered in the mission-orientation for knowledge production.

REFERENCES:

- Aagaard, K., Mongeon, P., Ramos-Vielba, I., & Thomas, D. A. (2021). 'Getting to the bottom of research funding: Acknowledging the complexity of funding dynamics', PLOS ONE, 16/5: e0251488.
- Braun, D. (2003). 'Lasting Tensions in Research Policy-Making - a Delegation Problem', Science and Public Policy, 30/5: 309–21.
- Eum, S. (2022). Project-Based Funding for Green Energy Technology by the Korean Green Mission and Its Changes and Characteristics." In 26th International Conference on Science, Technology and Innovation Indicators (STI 2022), Granada.
- —. (2023). Thematic research funding and the entry of researchers into prioritized areas: the case of green energy technology in Korea [preprint]. 27th International Conference on Science, Technology and Innovation Indicators (STI 2023).
- Fujimura, J. H. (1987). 'Constructing `Do-able' Problems in Cancer Research: Articulating Alignment', Social Studies of Science, 17/2: 257–93.
- Korean Government Ministries Concerned. (2009). 'Comprehensive Plan for the Green Technology R&D'
- Larrue, P. (2021). 'The Design and Implementation of Mission-Oriented Innovation Policies: A New Systemic Policy Approach to Address Societal Challenges', OECD.
- Laudel, G. (2023). 'Researchers' responses to their funding situation'. Handbook of Public Funding of Research, pp. 261–78. Edward Elgar Publishing. van der Meulen, B. (1998). 'Science policies as principal–agent games', Research Policy, 27/4: 397–414.
- Presidential Committee on Green Growth. (2009). 'National Strategy for the Green Growth'
- Shove, E. (2003). 'Principals, agents and research programmes', Science and Public Policy, 30/5: 371–81.

- Whitley, R. (2007). ‘Changing Governance of the Public Sciences: The Consequences of Establishing Research Evaluation Systems for Knowledge Production in DifferentCountries and Scientific Fields’. The Changing Governance of the Sciences: The Advent of Research Evaluation Systems, pp. 3–27. Springer Netherlands.
- —. (2014). ‘How Do Institutional Changes Affect Scientific Innovations? The Effects of Shifts in Authority Relationships, Protected Space, and Flexibility’. Organizational Transformation and Scientific Change: The Impact of Institutional Restructuring on Universities and Intellectual Innovation, pp. 367–406. Emerald Group Publishing Limited.

Barbara Glinsner, Dorothea Sturn, Anton Geyer
Centre for Social Innovation, Austria
Centre for Social Innovation , Austria
Inspire Research, Austria

THE TRANSFORMATIVE POTENTIAL OF IMPACT-DRIVEN INNOVATION

ABSTRACT

The programme "Impact Innovation" supporting non-traditional and social innovation Following two pilot calls in 2017 and 2018, the Impact Innovation programme was established as an ongoing call within the Austrian Research Promotion Agency (FFG). One of its main objectives is to fund innovation that creates a social, environmental or economic impact for a given target group and potentially beyond. In contrast to most of the other R&I funding programmes of the Austrian Research Promotion Agency, Impact Innovation employs a broader definition of innovation including process, service and social innovation. Through this broadened conceptualisation of innovation, the FFG also pursed the goal of enlarging access to FFG funding and thus attracting new applicants. A large proportion of the submitting organisations have no previous experience with the FFG and ist programmes. In addition to traditional corporations, there are also social enterprises, associations and sole proprietorships, often with social or educational policy concerns.

The third objective of the Impact Innovation programme is create more successful innovation projects by firstly employing innovation methods (e.g. Design Thinking, Open Innovation,....) and secondly by involving a diverse set of stakeholders, including potential target groups into their innovation processes.

The FFG commissioned ZSI and inspire research with an encompassing evaluation of the programme since 2019. The evaluation was conducted between June 2022 and December 2023.

The basic conceptual idea for the evaluation was to apply the innovative methods requested by the Impact Innovation programme to the evaluation itself. Accordingly, the topics and questions were addressed as part of a participatory process between the evaluators, the FFG, the representatives of the projects and various stakeholders. This process structured the accompanying evaluation in particular, while the results of the ex-post evaluation were used as input for reflection. The latter included an in-depth analysis of the documentation and data that were produced and/or collected by the FFG; an online-survey for successful(funded) and unsuccessful applicants that enabled control-group comparison as well interviews with the programme management and other experts within the FFG to investigate organisational learning processes triggered by the

programme that novel innovation objectives, processes and new innovators. The accompanying evaluation encompassed three online co-creation workshops with project managers of then ongoing Impact Innovation projects that focussed on the key elements of the programme (stakeholder participation and innovation methods) as well as on the theory of change and envisioned impact of the projects. In addation, a focus group with diverse representatives of the Austrian R&Isystem as well as of specific innovation communities (e.g. Start-Ups, Social Enterprises, Health Innovators,...) was conducted. The results of both elements were then validated and further discussed with FFG representatives in a final workshop.

The evaluation shows the positive perception of the special features of the programme, both on the part of the participating projects and on the part of central stakeholders. The key unique selling points of the programme are its openness to results, the early involvement of key stakeholders and the use of innovation methods. According to the funding recipients, the required approach to ImpactInnovation projects contributes significantly to the success of the projects: Six out of ten funding recipients stated that the flexible project design and the involvement of external stakeholders made a very large contribution to achieving the project goals.

Both the ex-post and the accompanying evaluation show that the projects took a very ambitious approach and involved both many and various stakeholders. The project managers described this involvement as purposeful and valuable. The processes were designed to be problem-centred and open to solutions, which was also seen as positive, albeit time-consuming. This appears to be one of the greater challenges for the projects: integrating busy external stakeholders in a goal-orientated and yet resource-saving manner in the short project duration. Accordingly, some of the funded projects expressed the desire for longer project durations in the context of the accompanying evaluation.

The projects were ambitious and willing to learn when it came to applying innovation methods. The concrete application of (innovation) methods enabled methodological knowledge and methodological expertise to be built up or expanded within the organisations, which will subsequently be used in further innovation projects. In the accompanying evaluation, design thinking emerged as the dominant innovation management method used. This raised the question of whether design thinking is possibly a guarantee that the innovation process will be designed as required by the programme.

In the ex-post survey, a very high proportion of funding recipients stated that the objectives of the projects had been achieved, with the required approach for Impact

Innovation projects contributing significantly to the success of the projects. Nevertheless, both ongoing and completed projects express the desire for a follow-up project or follow-up funding. This is not a contradiction when taking in consideration that the goal of an Impact Innovation project is achieved when a successful solution to the problem formulated at the beginning has been found. However, before this solution can be successfully implemented, further steps are required for which there is usually no space or time in the project itself.

Contribution of Impact Innovation to the transformation of socio-technical transformation: The programme Impact Innovation differs from traditional R&I programmes as it aims at broadening the innovation base by funding impact-driven process, service and social innovation as well as attracting new innovators and calls for participatory innovation processes that are driven by a sound methodology. This raises the question, whether this novel approach to funding innovation can contribute to socio-technical transformation?

The following evaluation results point in this direction. Firstly, it was evident from the ex-post evaluation as well as from the accompanying evaluation that eco-social rather than technical objectives are at the core of Impact Innovation projects.

Secondly, impact innovation increases of the innovation capacities of non-traditional innovators, particularly social innovators and social enterprises as well as organisations with a distinct socioecological orientation. This „new generation of innovators“ is actively seeking to find solutions to societal challenges and thus highly important in shaping the socio-technical transformation.

Thirdly, the idea of co-creating innovation and the integration of stakeholders and target groups into innovation processes ensures that different perspectives and voices are considered. When applied properly this cannot only make innovations more adopted to the needs of stakeholders, but can also contribute creating fairer and more inclusive innovation approaches.

However, there are also certain limitations to the contribution of Impact Innovation to the sociotechnical transformation. Firstly, the programme is comparably small in size, both from the overall budget as well as from the budget per project. Secondly, as with many policies or programmes supporting social innovation it can be said, that (funded) innovations are not disruptive per se, challenge or overturn current systems, but rather contribute to counterbalance negative impact of given system.

While the participatory approach of the evaluation design and the co-creative elements employed contributed to gaining a deeper understanding of the perspectives of project managers, the evaluators were faced with the same challenges that project managers of Impact Innovation projects faced: the limited time resources of participants and stakeholders. Participatory approaches therefore need be designed in way that they are not a burden on potential participants. Online workshops and facilitation methods proved to be quite time-effective, especially when inviting stakeholders from all over the country. This points to the question, how online methods and skills will gain further relevance in the future and which other participatory approaches or tools can be developed and employed that allow for deep insights while also being time-effective.

Lasse Bundgaard, Mireille Matt, Matthias Weber
Austrian Institute of Technology, Austria
LISIS INRAE, France
Austrian Institute of Technology, Austria

COMPARING TURNING POINTS: A TOOL TO ASSESS URBAN TRANSFORMATIONS

ABSTRACT

This paper investigates the relationship between Transformative Innovation Policies and turning points towards system innovation. TIPs are emerging at all levels of government; yet little is still known about how implementation impacts system level innovation. This is especially relevant at the local level, as initiatives like the Commission's City Mission are propelling – at least in terms of ambition – cities to the forefront of transformation by aiming to be climate neutral by 2030. At REvaluation21, we presented the first version of a framework, which sought to capture TIPs at the local level through five separate layers; (I) Guiding Visions, (II) Policy Logics, (III) Implementation Structures, (IV) Experimental Spaces, (V) Transformative Generalization. In this paper we revisit the framework above, applying it to four cases in order to disentangle an analysis of their sustainability trajectories as they have evolved over time. A comparative analysis will then form the basis from which we identify potential turning points in their trajectories towards system innovation.

Our initial framework was built from a literature review of innovation policy and urban studies. Combining these two literatures provided us with two idealized paths or modes of governance: transformative and planning. The former summarizes the theoretical expectations of TIP and Urban Climate Governance scholars, when it comes to a policy trajectory aimed at system innovation. The latter summarizes the two first generations of Innovation Policy (Science & Technology Policy and Innovation Systems Policy) and New Urban Governance scholars' expectation of how to use innovation policy to optimize existing systems (see elaboration below)

TABLE 1: FROM BUNDGAARD ET AL. (UNDER REVIEW)

Transformative IP/ Urban Climate Governance	Innovation Policy/ Smart Urban Governance
I. Guiding Visions	
Co-created Open to contestation Tool for coordination Not merely an add-on	Measurable Democratic through quantified data Top-down
II. Logics Policy	
System innovation Cross-cutting Learning-oriented	System optimization One policy, one instrument Efficiency
III. Implementation Structures	
New actors Bottom-up implementation Still a question mark in the literature	Usual structures Public-Private Partnerships Top-down Sectoral implementation
IV. Experimental Spaces	
Socio-technical or governance Inclusive	Strategic Urban space as a living lab for techno-logical experimentation Business development oriented
V. Generalization Processes	
Translocal diffusion Socially embedding & Shaping Multilevel interaction Evolutionary, radical, social, technological	Evolutionary process Continued big data collection enabling *scientific* scale up

Schot and Steinmueller (2018) define transformative change as situations where skills, infrastructures, organizational and industry structures, regulations, and preferences evolve, necessitating deliberate and coordinated strategic actions to achieve it. A turning point, therefore, facilitates a complete system reconfiguration, allowing a unified direction among various actors to form. Within transition studies both incremental, evolutionary changes and 'game-changing' moments can lead to system-level changes. However, our knowledge about what facilitates these turning points is still limited. Turning points are inherently linked to trajectories, with trajectories representing the continuation of a certain path and turning points being disruptions in this trajectory (Abbott, 2001). Consequently, understanding turning points require establishing trajectories, and they can only be identified by distinguishing between two different narrative trajectories separated by a point in time. Using the framework developed in our initial paper, we aim to identify trajectories within cities and, by relying on enablers identified in the literature, this paper seeks to understand what enables different types of turning points, whether transformative or non-transformative. In our first paper we used turning points as a heuristic tool to separate the analysis into different sequences. However, with this paper we aim to establish the theoretical underpinnings of turning points and their enablers within our five layers.

As a starting point, we rely on Loconto et al. (forthcoming) who identify enablers of turning points. As opposed to Loconto et al. we are not focusing on intermediation in the food sector alone, but rather a well-established and dense policy environment in cities, we add Borras et al. (2023) findings from a literature review on the transformative capacity of Public Sector Organizations. Each of the four enablers listed below represent an enabler for turning points. Turning Points are sufficient for transformative change, but transformation is not necessarily implied from a turning point. Ultimately this framework is not about assessing the transformative capacity of the four municipalities involved, but rather uses the insights from this concept to assess if and how transformative capacity triggers system innovation.

Knowledge & Resources
- Assets internal/external incl. natural, human, financial, legitimacy, mandate, network - Access to and communication of knowledge can minimize uncertainty and help shape the direction of a new path.
Values
- Agents of change wish to create "value" through system innovation and in this pursuit rally coalitions of actors behind new types of value to receive support and deepen the transition. By creating interdependencies between actors, new terms and concepts of value reduce uncertainty and accelerate turning points.
Skills & Infrastructures
- Anticipation, foresight, reflexivity, analytical, operational, coordination, adaptation and learning skills - Changes in organizational, network or physical infrastructures may lead to a turning point
Rules & Roles
- The roles played by PSOs influence the ability to change institutions, understood as formal and informal rules. - These roles can either maintain, create or disrupt institutions and through these tools initiate and uphold a turning point.

The comparative approach allows us to look at municipalities from Sweden and Austria that are all part of the Commissions NetZero Mission by 2030, and allows us to understand how turning points are affected differently through differences in the actors involved and the actual implementation of transformative policies. Going beyond case-studies, which has been the main method used in investigations of Transformative Innovation Policies, as well as focusing on implementation, rather than agenda-setting, this paper responds to calls in the literature for a diversity in methods and focus of research on TIPs (Haddad et al., 2022).

This contribution is relevant for the 1st thematic strand on Evaluation of transformation policies and dynamics in socio-technical systems. It is specifically relevant for the areas of "Monitoring national implementation of EU Missions", as the municipalities involved are addressing such an implementation albeit on a local and regional scale. It is furthermore relevant for "Assessing transformation policies", since the framework is meant as a methodology to assess the contribution of various transformative policies to turning points, be they transformative or not.

REFERENCES:

- Abbott, A. (2001). Time Matters: On Theory and Method. University of Chicago Press.
- Borr·s, S., Haakonsson, S. J., Poulsen, R. T., Pallesen, T., Hendriksen, C., Somavilla, L., ... & Gerli, F. (2023). The Transformative Capacity of Public Sector Organizations in Sustainability Transitions: A Conceptualization. Centre for Innovation Research (CIRCLE), Lund University.
- Bundgaard, K., et al. (under review). Developing a Conceptual Framework for Transformative Change at the City-level. Urban Studies and Innovation Policy.
- Haddad, C. R., Nakić, V., Bergek, A., & Hellsmark, H. (2022). Transformative innovation policy: A systematic review. Environmental Innovation and Societal Transitions, 43, 14-40.
- Loconto, A., et al. (under review). Intermediating Turning Points for Transformative Change. Journal of Environmental Policy and Planning.
- Schot, J., & Steinmueller, W. E. (2018). Three Frames for Innovation Policy: R&D, Systems of Innovation and Transformative Change. Research Policy, 47(9), 1554-1567.

Magdalena Wicher
University of Bergen (UIB), Vienna Science & Technology Fund (WWTF), Institute for Advanced Studies (IHS), Austria/Norway

MOVING BEYOND CYCLES OF FUND AND FORGET – RESEARCH AND INNOVATION GOVERNANCE BRIDGING THE GAPS IN TRANSLATION FROM POLICY DESIGN TO PRACTICE

ABSTRACT

Object and purpose of the presented approach Changing demands on the production of knowledge are accompanied with changes and challenges for the science system. There are different reactions and mechanisms how to respond to these changes. Mission-orientated policymaking in research and innovation (R&I) has attracted increasing attention in recent years. These types of policies are linked to major societal challenges such as climate change, an ageing society, migration and inequality. Multi-, interdisciplinary and transdisciplinary approaches and the involvement of society should lead to an opening of science that is orientated towards societal and ecological demands and supposed to support socio-technical transformation. In this vein, Transformative Innovation Policies (TIPs) take a similar direction.

This kind of R&I policy thinking has in common that they are based on an idea of knowledge production that goes beyond traditional ideals (see Mode 2 research, Gibbons et al. 1994). Schot & Steinm¸ller (2018) offer a suitable model with their ¥third frame¥ of R&I policy, which extends the policy focus beyond inputs and actors to the ¥outcomes¥ of policy interventions and attempts to direct R&I towards environmental and social challenges. Through experimentation and expansion, the idea is that transformative innovation policies promote dynamics that drive long-term changes in existing conditions.

Changes in the science system in recent decades have also brought with them a commercialization that is reflected in directive, short-term project thinking. New Public Management (NPM) systems counteract traditional models of academic self-governance (Sandstrˆm & Besselaar, 2018), also by introducing refined evaluation and monitoring procedures and indicator assessment (Braun 2003). Projectification" (Torka, 2006) as a management approach has become an important form of research governance, leading to practices that prevent policy, funding or implementation from thinking and acting in the long term. The dynamic of ¥Fund and Forget¥ seems to make it difficult to bring about long-term learning, upscaling of outcomes and change. Other forms of governance are needed that take particular account of the role of policy and funding in the types of knowledge production required.

Novelty (policy design, approach, methodology): The aim of this presentation is not to reopen old problems in the R&I system, but to put forward provocative and radical proposals for change and discuss them with the audience. The focus is on the question of how to get out of the ¥Fund and Forget¥ dynamic. The goal of this contribution is to identify the similarities and differences between different concepts of research, technology and innovation policy, which can be summarized under rubrics including mission-oriented policies (Mazzucato 2017), grand challenges, third-generation innovation policies, and transformative innovation policies (Diercks et al. 2019; Haddad et al. 2022; Schot and Steinmueller 2018). Therefore, this paper aims to analyze findings from previous measures with similar characteristics and their potential for a transformative way of producing knowledge in the science system. Different examples will be used to illustrate how policy ideas and approaches are implemented and translated into practice and what can be learnt from them.

To this end, the concept of Responsible Research and Innovation (RRI) is used as one case study for the attempt to steer research and innovation. It will show how RRI was presented in the EU Framework Programme SWaFS (Science With and For Society) and translated and implemented in individual projects. Another example is a transnational European funding Call focusing on transformative research in the face of climate change. Initiated by JPI-Climate (Joint Programming Initiative "Connecting Climate Knowledge for Europe"), the topic "Enabling Societal Transformation in the Face of Climate Change" was funded as part of the SOLSTICE Call where Social Science and Humanities (SSH) projects were to conduct research on climate change.

While RRI emerged as a promising approach to redefining the relationship between science and society, the concept as a whole – though still present in individual aspects – has not gained traction in the European policy and research funding landscape. This development makes the case a particularly interesting one and encourages an examination of the underlying governance mechanisms, agendas and objectives that have steered the discourse. The analysis will show where and how gaps between policy idea, funding and implementation have occurred.

SOLSTICE, in turn, demonstrates the essential potential of transformative innovation policies with its innovative approach of taking SSH into the leadership role. Ambitious policy ideas for societal transformation were translated into a funding call, followed up with seven funded projects. Interdisciplinarity played a key role in implementation, whereas transdisciplinarity and the involvement of non-scientific stakeholders lost their potential due to structural funding hurdles. Potentials, weaknesses and barriers in the translation of theory into practice can also be clearly identified here.

Even if the examples start at different levels, they do have several aspects in common: demandorientation (Boon and Edler 2018; Grillitsch et al. 2019 refer to ¥demand articulation¥; Serger and Palmberg 2022; Smits and Kuhlmann 2004) and the resulting directionality of R&I policymaking
(Aagaard et al. 2022; Borr·s and Schwaag Serger 2022; Diercks et al. 2019; Grillitsch et al. 2019; Haddad et al. 2022; Lindner et al. 2016; Schot and Steinmueller 2018; Serger and Palmberg 2022); challenges arising in complex systems and the search for new paths and ways to solve them (Ghosh et al. 2021; Howoldt and Borr·s 2023); interdisciplinarity; inclusive and participatory processes and multi-actor implementation (Borr·s and Schwaag Serger 2022; Ghosh et al. 2021; Haddad et al. 2022; Howoldt and Borr·s 2023; Kuhlmann and Rip 2018); an essential component is also the will to learn (deep / second-order learning), reflexivity and reflection throughout the entire process (Ghosh et al. 2021; Grillitsch et al. 2019; Haddad et al. 2022; Molas-Gallart et al. 2021; Serger and Palmberg 2022; Smits and Kuhlmann 2004).
Results (if already available): In both case studies, the ¥Fund and Forget¥ dynamic seems to prevail. The presentation will provide insights into the similarities, take into account the different backgrounds and analyze the dynamics of implementation. For example, joint strategies for upscaling the outcomes at project and programme level are negotiated differently in the examples presented and are mostly missing. These strategies could help to share the outcomes and lessons learnt and then also to carry them forward to initiate change. The contribution is intended to provide further insights into possibilities and suggestions as to how these dynamics and governance mechanisms can be accommodated.
Finally, it should be emphasized that we, as evaluators, could see our role throughout the process as exploring the possibilities of transformative approaches. We could act as an additional catalyst for the transformative potential of policy and funding (and their future iterations) and also funded projects. In this way, we could help bridge the gap between funders, researchers and stakeholders concerned with the theory and practice of transformative change.

Maria-Ascensión Barajas, Marina Onetti, Stefanie Schuerz
Centre for Technological Development and Innovation (CDTI) , Spain
Andaira Coop, Spain
ZSI, Austria

EVALUATING SOCIAL IMPACT OF R&D PUBLIC PROGRAMMES: A PARTICIPATORY APPROACH

ABSTRACT

Even though science and technological progress is undoubtedly positive for society, there is a lack of evidence on the real social impact of R&D and innovation public programmes. Key questions arise: Are innovations equally accessible to citizens? Do the results of supported R&D projects solve real social problems? Do R&D programmes contribute to the sustainability of the productive models? Is there any collective harmed by R&D programmes? In this regard, some social issues are considered as eligibility criteria (for instance, mandatory gender equality plans for companies or DNSH normative) but this procedure is not a valid tool for assessing the impact of public programmes on society. Policymakers should base their decisions on evidence, and the goal of policy evaluation is to provide this evidence. The problem is that, traditionally, impact evaluation of R&D and innovation programmes has been focused on technological and economic results, but not on the social effects. The challenge to improve R&D policies is to shift the focus from a pure technological approach to a social one.

The political sphere of an organisation should not be considered a boundary for the management of public programmes; it is necessary to incorporate a broader approach in order to understand the repercussions that these programmes have on society. Evaluation is a tool to open up this approach and generate knowledge for decision-making from a participatory point of view, opening up the organisation to society. It is necessary to make the social links of R&D policies visible so that these policies really contribute to improving people's lives.

Being aware of this challenge, the CDTI decided to open its evaluative approach towards social issues and carry out a pilot project with the Neotec programme. Neotec provides grants for New Technology-Based Firms which carry out projects that require the use of technologies or knowledge developed from research activities and with a viable business plan. As well as contributing to entrepreneurship, Neotec is also designed as a tool to promote knowledge transfer from public research bodies to society. Due to its multiple social implications, Neotec was considered a good pilot to test new open evaluation approaches based on citizens' participation and social impact.

The criteria for evaluating applications prioritise technological excellence (35% of the total score), the viability of the business plan (30%), and the management capacity of the

entrepreneurial team (30%). Social impact criteria are present but with a very low weight (5%). The programme targets start-ups up to three years of age and technology-based business plans with a low level of maturity at the time of application. This means that the level of risk is high, from the point of view of both business development and technology.

In the 2018-2020 period, the annual budget of the programme was €25 million. From the last year onwards, the budget increased to reach €40 million in 2023. This evaluation has taken into account the calls for proposals for the years 2018 to 2022, during which 550 projects were approved, representing 25% of the applications submitted.

DEFINING SOCIAL IMPACT THROUGH A PARTICIPATORY APPROACH: The participatory approach will allow to define from an open point of view the concept "social impact" and identify those stakeholders who should be taken into account to define the effects of NEOTEC on the society. At this regards, ethical issues are key aspects, not just because there are legal issues to deal with, but also because people engaged in participatory activities usually do not belong to the institution and are not familiarized with internal procedures. In this context. they should be guided and informed within a valid ethical framework.

The relevant stakeholders were assigned to different groups defined in the Stakeholders Circles Model, according to their proximity to Neotec, and their role in society: circle of control (beneficiary and non-beneficiary companies; programme managers), circle of influence (regional governments; entrepreneurship centres; consulting companies) and circle of interest (university foundations; gender equality entities and NGOs).

Stakeholders were invited to participate in the evaluation of the social impact of Neotec at three different stages: definition of the evaluation questions; information collection and interpretation of results.

The participatory activities carried out during the evaluation have taken into account the outputs of the European PRO-Ethics project, funded through the Horizon 2020 programme. The Ethic Framework and the Guidelines facilitate tools for the implementation and management of ethical aspects in participatory processes carried out by public innovation agencies.

The most important collaborators were the participants in the first phase of the evaluation (beneficiary companies and consultants; civil society organisations such as ONCE, Innovatia 8.3 or ANCES; and governmental organisations such as CDTI and regional innovation agencies), who acted as informants, helping to shape the definition of the social impact to be measured and giving the evaluation its participatory character, adapted to the interests of the participants themselves.

The time frame for this evaluation was 2018-2022. The information was extracted from a representative sample of 614 companies with a Neotec project approved or rejected during this period which answered a survey sent to 1.317 firms. A participatory approach has been applied and quantitative (PSM, differences in differences) and qualitative (focus group, interviews) methodologies have been used.

The main challenges faced by the evaluation have been: identifying the relevant actors who had knowledge of the programme environment and vision of the social impact of technology; the programme was not designed to address social issues, it is an area "outside" the public intervention that was being assessed; the cultural and organisational barriers to consider the social dimension, especially in the R&D environment; less involvement of the actors furthest away from the programme due to lack of interest or motivation as they do not see the benefits of participating in the evaluation; and creating links with actors linked to social issues, as the CDTI's activity is not perceived as relevant to solving social issues.

In order to incorporate social issues in R&D policies it is crucial to develop a culture of evaluation in public administration and in society at large. Trained teams and financial resources are needed for this purpose. In the same way, initiatives such as PRO-Ethics or applying a social perspective to R&D policies are examples of best practices that bring added value to society and that should be taken into account in order to consolidate progress towards the evaluation of public policies.

Furthermore, it requires the support of the organisation's management team and alignment with the policy priorities of the responsible ministry; and human resources with training in evaluation methodology and motivation to introduce a change of perspective.

Replicability of the innovation is ensured through the systematic use of a robust methodology used in the evaluation. This means that the same procedures can be applied to conduct similar evaluations, ensuring that the social impact of the various programmes implemented in R&D policies are measured in a reliable way.

The Neotec programme is having a positive impact on knowledge generation. There was a clear and consistent impact on the funded companies for which the grant was not awarded: •

- They have seen a greater increase in R&D&I expenditure and "intangible" fixed assets.
- They have more lines of R&D&I and further-reaching strategies.
- They register more patents.

These aspects are, in turn, linked to the positive impact that Neotec has on the transfer of technology, particularly from the university, whether through the use of technology patented by a research team, the hiring of research personnel by companies or collaboration between these personnel and their original organisation.

Moreover, the programme has limited or partial results in terms of social and environmental impact. In particular:

- Neotec does not have a positive impact on technology development processes with a gender perspective or on the internal business management activities associated with the work team. In this regard, there is no corrective effect of the programme on the presence of women in the different levels of the organisation, including the positions that hold share capital.
- Neotec has no differential impact on the development of technology aimed at solving social problems, although a high percentage of projects (60%) are related to health and quality of life. Moreover, it does not help minimise the technology gap of groups at risk, and there is even a negative effect on access to technology by some groups, such as people with low income levels or those living in rural areas.
- The effects on job creation for social groups with greater difficulties in accessing the labour market are partial, highlighting a positive effect on young people under the age of 25.
- The programme has no different impact on the development of sustainable technologies, and is even negative in the area of adaptation to climate change and the sustainable use and protection of water resources. This means that the programme alone fails to boost these technologies.

These weak or partial results in terms of socio-environmental impact can be explained by a lack of market motivation or stimuli from companies, as well as a lack of programmatic emphasis on these aspects: the criteria of social impact in the Neotec calls have a very little impact in the awarding of the grant (5 points out of 100 as of the 2019 call) and, given their multi-dimensional nature, present difficulties in scoring the applications.

The beneficiaries of the evaluative process are, on the one hand, the collaborators themselves, as this evaluation influenced their decision-making. On the other hand, the community of public policy evaluators also benefits from this evaluation because of its innovative character, as it broadens the knowledge on how to do evaluation, especially in the R&D field. Policy makers and R&D promotion authorities get recommendations to guide their policies, ultimately benefiting society as a whole.

This evaluation becomes a learning strategy developed by CDTI, which allows the social impact to be incorporated into other evaluations using well-defined criteria.

It also opens a reflection on the evaluation criteria of the proposals submitted, leading to a review of the social aspects that are taken into account and the total weight they have in the score of the proposals. In fact, other regional agencies expressed their intention to incorporate social criteria in their own processes.

Actors in charge of the evaluation and development of public R&D policies can benefit from the evaluation results, both in evaluating their own programmes and in developing new ones.

The methodology for social impact assessment has been definitively incorporated into CDTI's evaluation strategy. The foundations have been laid for a shift towards the consideration of social impact as an integral part of R&D support programmes.

Julia Melkers, Nicolas Robinson-Garcia, Luyu Du
Center for Organization Research and Design (CORD), Arizona State University, United States
Universidad de Granada, Spain
Center for Organization Research and Design (CORD), Arizona State University, United States

THEY COLLABORATE, BUT DO THEY "CO-PRODUCE?" EXAMINING ACADEMIC SCIENTISTS' COLLABORATIVE PATTERNS WITH NONACADEMICS

ABSTRACT

The funding and conduct of science is aimed at knowledge advancement but also in improving the human condition (Polanyi, 2000). Policymakers and the general public expect that new knowledge discoveries derived from scientific and engineering research, particularly publicly-funded research, will diffuse and improve society (Sarewitz, 2016). Aligning academic research with societal needs has become an increasingly pressing concern, especially when considering the substantial public funding in science (Crow & Dabars, 2015; Gerber et al., 2023). These issues are not limited to the U.S. context, globally questions of how to enhance societal benefits for research (de Jong et al., 2016), and in turn, how to measure it (Bornmann, 2013), are growing in the research policy community. We may safely conclude that attention to and expectations for broader impacts from the scientific community are here to stay. What is less clear, is the capacity of academic researchers to be in the position to contribute to broader impacts, in particular to diffuse knowledge to non-academic communities, and how this varies across the scientific community.

While there are a number of mechanisms to accomplish broader impacts, we examine faculty engagement in co-production processes with non-academic partners. Given the persistent knowledge-action gap, co-production of knowledge emerges as one of the most important ideas among scholarly discourses (Miller & Wyborn, 2020). We are specifically interested in knowledge co-production, which involves meaningful collaboration integrating diverse perspectives of academics and non-academics throughout the research process, which is essential for applying scientific knowledge to understand and address complex social problems (Yua et al., 2022). Research on these types of linkages has shown that cross-sector collaboration transfers knowledge in the process (Spaapen and Drooge, 2011), and also tends to produce wellcited journal articles (Lebeau et al., 2008), suggesting even broader (although not necessarily societal) impacts. Researchers who collaborate across sectors often respond to societal demands on the production of new knowledge. Considering both academic and non-academic perspectives in a co-production process increases the potential of research to have real

impacts (Trencher et al., 2014).

In light of increasing pressure to demonstrate societal impacts of research, understanding what drives academics to engage with non-academic partners is critical to supporting incentive and reward structures in universities that support these alliances. Much of the work in academic collaboration with other sectors has focused on university-industry relations (important for commercialization) or public research agencies/laboratories (involving collaboration among scientists in both sectors). Yet, co-production can also involve engagement with a range of nonacademic organizations, both research-oriented and non-research oriented. Examples include collaboration with non-profit or local government organizations around environmental threats, education research with local school districts, or emergency management mitigation, among many others.

In this paper, we ask: what are the characteristics of faculty who collaborate with nonacademics, particularly those in non-research-oriented organizations? Drawing from the scientific & technical human capital (STHC) theory, we examine the human and social factors that matter in the successful collaboration with these non-academic partners. We examine both the inclusion of non-academic partners in close collaborative networks, and those with whom academics successfully publish. We recognize that neither of these reflect the diverse and often complex dimensions and outcomes of co-production. However, they do reflect the important relational aspects, and those that focus on products (in this case co-authorship) as a result of those relations. We also extend the body of literature focused on the research-centric triple helix collaborative patterns of academic science (Etzkowitz and Leydesdorff, 1998), to consider a broader range of sectoral types with whom academics collaborate, both research and nonresearch based institutions.

The STHC model focuses on understanding the development of researchers' capacities of know-how through their technical knowledge and training, professional networks, and resources (Bozeman et al., 2001; Corley et al., 2019). It posits that academic scientists do not just accumulate knowledge and skills in isolation; instead, they build their human and social capital through interactions and collaborations not only within their institutional environments but also in broader professional networks (Ponomariov & Boardman, 2010). Our study also has the potential to extend the STHC model by delving into the diverse collaborative patterns between academics and non-academics—which is an important yet underexplored indicator of scientific and technical performance.

We hypothesize that:

H1a: Faculty who have had broad career experience outside of academia will be more likely to develop close professional network ties with non-academics.

H1b: Faculty who have had broad career experience outside of academia will be more likely to collaborate with non-academics.

H2: Faculty with greater career security and resources will be more likely to collaborate with non-research organizations than those with less career security and resources.

Data Sources: Our research is based on data drawn from the NETWISE II study and included survey, CV and lifetime bibliometric data for each survey respondent (citations removed). These data are relatively unique because they include a national sample of academic scientists with detailed demographic and work history data that is matched to bibliometric records, enabling extensive analysis of publication data that is not possible without these additional individual data. The core of the study (implemented in 2012) involved an extensive survey of tenured/tenure-track STEM faculty in the United States in four disciplinary areas: biology, biochemistry, civil engineering and mathematics, and the collection of respondent CVs. Using the 2000 Carnegie Classification system, the sample included Research Extensive and Research Intensive institutions, as well as teaching-centric institutions (Master's I/II and elite liberal arts colleges). These institutional types included here account for nearly 28% of all post-secondary institutions in the United States, and nearly 75% of all 4 year institutions. The survey sample was stratified across institutional type, faculty rank and discipline, and oversampled for gender, resulting in a final sample of 9,925 (38% of the original sampling frame). The survey was implemented online and had a total unweighted response rate of 42%, with 4,195 completed or partially completed surveys submitted. The survey addressed a broad set of items relevant to the study of academic careers in STEM. Given our interest on research collaboration for co-production, we focus on research-oriented universities, for a final sample of 2,111. Of these, we had CVs and were able to match bibliometric data for a final sample size of 1,593.

A strength of our study is that we have detailed survey and demographic data that is matched with lifetime bibliometric records drawn from the Web of Science, initially gathered in 2013 and then updated in 2019. The NETWISE II study also involved the collection of life-time publications for all survey respondents (through 2013 when the survey was completed) indexed in Web of Science (WoS). Publications were matched using an algorithm that linked publication title words' clusters in an individual's CV with the text of their bibliometric record. This involved using a boosted-trees method for name disambiguation (Wang et al., 2012) to each of the respondents, and checked through a detailed effort of alignment with respondents' CVs. To be conservative regarding this matching, data from respondents were included only if they had a valid CV and if publication data was identified. The final data were checked for bias in this matching process, with limited differences for African American faculty and faculty in Mathematics. Based on these inclusion criteria, our final dataset included 1,122 respondents from each of the institutional types and disciplines, with a total 9,044 publications. In our analysis, we rely on co-authorship data to identify co-production patterns across various institutional types. For the 9,044 publications of the 1,122 researchers, 20,580 unique affiliations were identified.

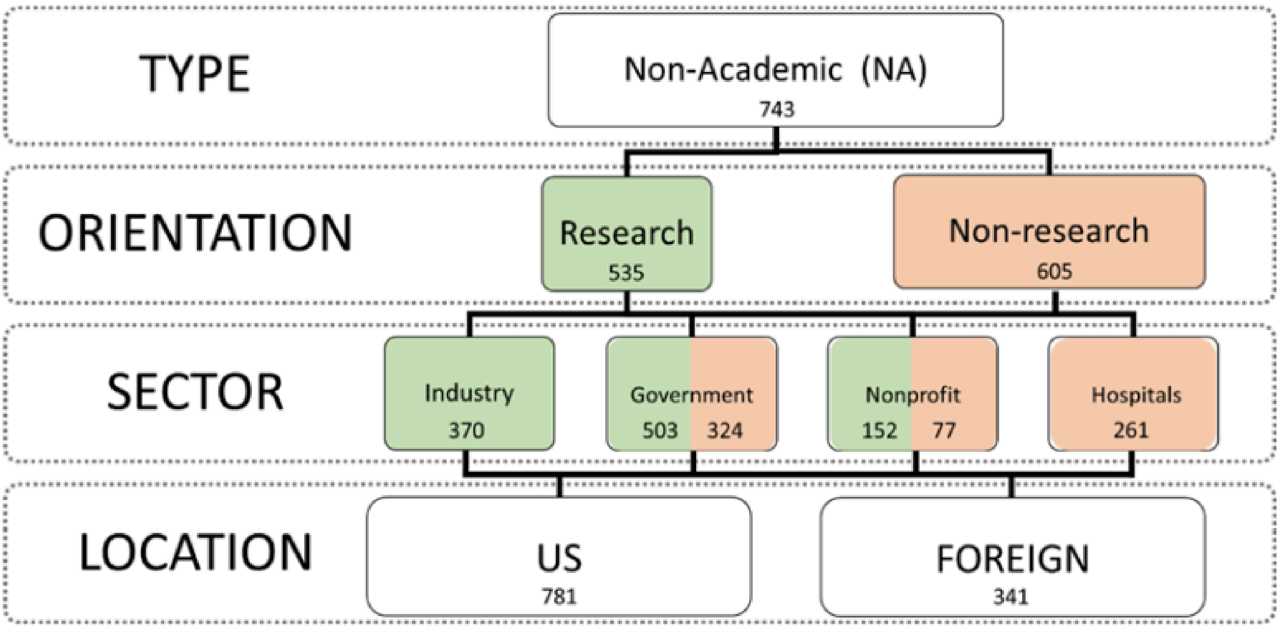
Operationalization:

TABLE 1 LISTS THE DEPENDENT VARIABLES FOR THE VARIOUS MODELS, AND THE INDEPENDENT VARIABLES (SAME) FOR EACH MODEL, ALONG WITH THEIR OPERATIONALIZATION.

Construct	Operationalization
Dependent Variables	
Non-academic Orientation	**Non-Academic Co-Authorship:** Co-authored with individuals in nonacademic organizations; research-oriented NA organizations; non-research-oriented organizations. **Research orientation**: Co-author(s) affiliated with an academic institution or a research oriented non-academic institution (e.g., governmental labs, nonprofit research centers) **Commercial Orientation:** Co-author(s) affiliated with industry **Societal Needs Orientation:** Co-author(s) affiliated with hospitals or clinical centers; Co-author(s) affiliated with non-academic non-research institution (e.g., government, nonprofits)
Non-academic research intensity	**E -I Index** for **a)** percentage of papers with non-academic and academic co-authored publications and **b)** with NA research and NA non-research.
Independent Variables	
Demographic characteristics	Respondent gender (M/F) and academic rank (Full, Associate, Assistant Professor); Citizenship; non-academic career experiences
Research Activities	Time on research; industry funding; center affiliation; grants activities
Institutional type	Categorized by the Carnegie Classification System (Research Extensive, Research Intensive)
Discipline	Disciplinary Academic Department: Biology, biochemistry, civil engineering and mathematics
Productivity	Industry funding received, number of research proposals submitted and proportion of time spent on research

To measure respondent collaborative affiliations, institutional data from scientists' publication records was manually coded by two independent team members into a sectors taxonomy to identify non-academic collaborators. When mismatches in coding were found, both experts would come into an agreement. To test for coding bias, each researcher coded ten percent of one another's codes. Tests of significant differences were conducted to determine whether there were differences in the application of the coding taxonomy (shown in Figure 1).

FIGURE 1. TAXONOMY OF NON-ACADEMIC (NA) SECTORS. NUMBER OF RESPONDENTS WITH ANY NA COLLABORATION.



First, institutions were coded based on their geographical location as foreign or US institutions. Academic institutions include universities, colleges or institutes of technology. Second, all non-degree granting institutions were coded as non-academic institutions (referred to as NA hereafter). Within the NA institutions four sectors were identified: industry (e.g., private firms, companies), government (e.g., governmental labs, state or provincial government, national ministries), nonprofits (e.g., independent research labs, international organizations, professional or scientific associations) and hospitals (including healthcare providers or private clinics).

Sectors were categorized based on the primary organizational mission (research or not) (Figure 1). Industry was coded as research, while government and nonprofits were coded depending on the type of institution, and hospitals were considered non-research oriented. For example, government labs or centers depending of national academies (e.g., Chinese Academy of Science or Max Plank) were considered as research organizations. Agencies dependent of national ministries, local councils, etc., were considered governmental non-research institutions. In the case of nonprofits, joint research centers and independent labs were identified as research oriented (e.g., Helmholtz Association of German Research Centres) while international organizations, zoos or museums were considered non-research oriented (e.g., Animal Wildlife International), given that this is not their primary mission. Dummy variables were constructed to indicate the existence of any of these collaborations, and the geographic location of each coauthor, for both U.S.-non-U.S. and different global regions defined by the World Bank classification (https://data.worldbank.org). Affiliation was coded as a binary variable if the respondent had co-authored a paper from each given sector.

To analyze the balance of collaboration across the academic and non-academic sectors (referred to here as NA Intensity), we calculated the E-I Index commonly used in social network studies (Krackhardt and Stern, 1988), which captures the balance of a node in network between its interactions within a cluster and its interactions with the rest of the network. It ranges from -1 to 1 where -1 is interpreted as having all ties with other nodes from the cluster (internal) while 1 means that the analyzed node has all ties with the rest of the network (external). Respondents with E-I indices greater than 0 and close to 1 would show a preference for NA collaboration.

Analysis and Preliminary Results: Our analysis involves both descriptive summaries of engagement and co-authorship with non-academic partners, including the variation in those types of organizations. We are also conducting a series of regressions (OLS and Logit) to examine the different ways that overall as well as broad career-related individual human capital matters for non-academic social capital, and the outcomes of those ties. While these models are still being adjusted (and not shown here), our preliminary results show that there are distinct patterns that reflect career stage and independence in the types of non-academic collaborations that U.S. faculty develop. Our descriptive results (Table 2) show that collaboration with non-academics is the norm across faculty in each discipline and institutional type, with about two-thirds of all respondents having collaborated with non-academics at least once. Further, with the exception of Mathematics faculty, only about one-third in any of these groups have collaborated with academic collaborators only.

TABLE 2. DISTRIBUTION OF RESPONDENTS BY COLLABORATION SECTOR

		Non-Academic collaboration			Research Oriented			Commercialization	Societal Values		
	TOTAL	NA	NAR	NANR	ACAD	GOVR	NPR	IND	HOSP	GOVNR	NPNR
Citizenship											
Foreign	341	67%	47%	56%	33%	44%	12%	36%	27%	32%	4%
US. Born	781	66%	48%	53%	34%	45%	14%	32%	22%	28%	8%
Gender and Rank											
Female	491	63%	45%	49%	37%	41%	15%	29%	23%	25%	7%
Male	631	69%	50%	58%	31%	48%	13%	36%	23%	32%	6%
Assistant	302	77%	53%	64%	36%	50%	33%	37%	27%	15%	8%
Associate	368	63%	44%	52%	37%	41%	13%	30%	22%	27%	6%
Full	445	70%	52%	59%	30%	50%	13%	40%	28%	32%	8%
Discipline											
Biology	453	72%	56%	59%	28%	52%	19%	30%	28%	35%	12%
Biochem	234	68%	51%	52%	32%	46%	15%	33%	29%	24%	2%
Civil Eng	202	80%	48%	73%	20%	47%	9%	59%	18%	41%	6%
Math	233	42%	29%	30%	58%	28%	6%	16%	13%	12%	2%
Institutional Type											
Research Int	217	65%	48%	53%	35%	42%	17%	32%	25%	30%	7%
Research Ext	402	79%	57%	66%	21%	55%	18%	44%	26%	37%	7%
Total	1,122	66%	48%	54%	34%	45%	14%	33%	23%	29%	7%
N	1,122	743	535	605	379	503	152	370	261	324	77

ACAD: academic authorship only, IND: industry co-authorship, GOVR: government research, GOVNR: government non-research, NPR: nonprofit research, NPNR: nonprofit non-research, HOSP: hospital

Disciplinarily, Civil Engineering has the highest shares with around 80% of respondents having co-authored with NA stakeholders, while the majority of Mathematics faculty have only collaborated with other academics (58%). Interestingly, results show that faculty in all institutional settings collaborate with non-academics, with the highest proportion in Research Extensive institutions. Interestingly, those working in the teaching-centric Liberal Arts institutions are equally distributed with half of them reporting NA collaborations. A slightly smaller proportion of women collaborate with NA stakeholders (63%) than do men (69%), and the largest differences exist by academic rank; 70% of full professors collaborate with NA sectors while 63% of associate professors and 77% of assistant professors do so.

Our interest extends beyond the simple designation of non-academic collaborators. When examined by mission (research versus non-research) and sector (industry, government, non-profit and hospitals), results show some variation across respondent characteristics. Further, proportionally, foreign-born scientists collaborate more with non-research non-academic organizations than do U.S. born scientists, which is unexpected given our initial hypothesis.

Almost half of respondents have collaborated with government research institutions (45%), while only a small subset of respondents have published with Nonprofit organizations, either with a research (14%) or a non-research focus (7%). Further, this pattern only diverges when disaggregating by fields. Here, civil engineers prefer collaborating with industry rather than government research, biochemists prefer collaborating most with government research institutions, followed by industry and hospitals and biologists tend to collaborate with government research institutions followed by government non-research and industry.

Overall, the results shown in Table 2 demonstrate that non-academic collaboration is common across all groups of academic scientists in all disciplines in our sample, and that this collaboration extends beyond research organizations to others that may provide the foundation for the dissemination of knowledge relevant to broader impacts.

Discussion and Conclusion: Our paper examines variations in the academic and non-academic collaborative preferences of academic researchers in the U.S. academic system. We are particularly interested in how collaboration that may enable translational research and stakeholder-relevant outcomes fits with the traditional academic paradigm. Partnering with different stakeholders in knowledge production is critical for socially relevant science (Ely et al., 2014; Spaapen and Drooge, 2011), yet, prior work has shed little light on the preferences and characteristics of academic researchers who engage outside of academia and across this broader set of stakeholders. One mechanism for this transfer is through active collaboration, operationalized here via co-authorship. While this is clearly a limited view of collaboration, it does reveal patterns in the relationships developed across the academic workforce, and how that varies by faculty characteristics. Further, we contribute to how these behaviors extend beyond industry and government labs, which have been focal points of prior research (Hayashi, 2003; Rahm et al., 1988). In our work, we focus on industry, but also a broader set of collaborative affiliations that include but go beyond researchcentric organizations.

REFERENCES:

- Ackers, L., 2005. Moving People and Knowledge: Scientific Mobility in the European Union1. Int. Migr. 43, 99– 131. https://doi.org/10.1111/j.1468-2435.2005.00343.x A
- garwal, R., Ohyama, A., 2012. Industry or Academia, Basic or Applied? Career Choices and Earnings Trajectories of Scientists. Manag. Sci. 59, 950–970. https://doi.org/10.1287/mnsc.1120.1582
- Altbach, P.G., Knight, J., 2007. The Internationalization of Higher Education: Motivations and Realities. J. Stud. Int. Educ. 11, 290–305. https://doi.org/10.1177/1028315307303542
- Andújar, I., Cañibano, C., Fernandez-Zubieta, A., 2015. International Stays Abroad, Collaborations and the Return of Spanish Researchers. Sci. Technol. Soc. 20, 322–348. https://doi.org/10.1177/0971721815597138
- Åslund, O., Hensvik, L., Skans, O.N., 2014. Seeking similarity: How immigrants and natives manage in the labor market. J. Labor Econ. 32, 405–441.
- Bang, H., 2015. Challenges and self-efficacy of female East Asian-born faculty in American universities. Exp. Immigr. Profr. Cross-Cult. Differ. Chall. Lessons Success 88–99.
- Besharov, M.L., Smith, W.K., 2014. Multiple Institutional Logics in Organizations: Explaining Their Varied Nature and Implications. Acad. Manage. Rev. 39, 364–381. https://doi.org/10.5465/amr.2011.0431
- Bonney, R., Cooper, C.B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K.V., Shirk, J., 2009. Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. BioScience 59, 977–984. https://doi.org/10.1525/bio.2009.59.11.9

- Bornmann, L., 2013. What is societal impact of research and how can it be assessed? a literature survey. J. Am. Soc. Inf. Sci. Technol. 64, 217–233. https://doi.org/10.1002/asi.22803
- Bozeman, B., Corley, E., 2004. Scientists' collaboration strategies: implications for scientific and technical human capital. Res. Policy, Scientific and Technical Human Capital: Science Careers and Networks as Knowledge Assets 33, 599–616. https://doi.org/10.1016/j.respol.2004.01.008
- Bozeman, B., Fay, D., Slade, C.P., 2013. Research collaboration in universities and academic entrepreneurship: thestate-of-the-art. J. Technol. Transf. 38, 1–67. https://doi.org/10.1007/s10961-012-9281-8
- Bozeman, B., Sarewitz, D., 2011. Public Value Mapping and Science Policy Evaluation. Minerva 49, 1–23. https://doi.org/10.1007/s11024-011-9161-7
- Breznitz, S.M., Feldman, M.P., 2010. The engaged university. J. Technol. Transf. 37, 139–157. https://doi.org/10.1007/s10961-010-9183-6
- Burggren, W.W., 2009. Implementation of the National Science Foundation's "Broader Impacts": Efficiency Considerations and Alternative Approaches. Soc. Epistemol. 23, 221–237. https://doi.org/10.1080/02691720903364092
- Butterfield, A.K.J., Soska, T., 2013. University-community partnerships: Universities in civic engagement. Routledge.
- Cañibano, C., Otamendi, J., Andújar, I., 2009. An assessment of selection processes among candidates for public research grants: the case of the Ramón y Cajal Programme in Spain. Res. Eval. 18, 153–161. https://doi.org/10.3152/095820209X444968
- Carson, L.R., 2009. "I am because we are:" collectivism as a foundational characteristic of African American college student identity and academic achievement. Soc. Psychol. Educ. 12, 327–344.
- Ciarli, T., Rafols, I., 2018. The Relation Between Research Priorities and Societal Demands: The Case of Rice (SSRN Scholarly Paper No. ID 3093285). Social Science Research Network, Rochester, NY.
- Collins, J.M., 2008. Coming to America: Challenges for Faculty Coming to United States' Universities. J. Geogr. High. Educ. 32, 179–188. https://doi.org/10.1080/03098260701731215
- Corley, E.A., Bozeman, B., Zhang, X., Tsai, C.-C., 2017. The expanded scientific and technical human capital model: the addition of a cultural dimension. J. Technol. Transf. 1–19.

- de Jong, S.P.L., Smit, J., van Drooge, L., 2016. Scientists' response to societal impact policies: A policy paradox. Sci. Public Policy 43, 102–114. https://doi.org/10.1093/scipol/scv023
- D'Este, P., Perkmann, M., 2011. Why do academics engage with industry? The entrepreneurial university and individual motivations. J. Technol. Transf. 36, 316–339. https://doi.org/10.1007/s10961-010-9153-z
- D'Este, P., Ramos-Vielba, I., Woolley, R., Amara, N., 2018. How do researchers generate scientific and societal impacts? Towards an analytical and operational framework. Sci. Public Policy. https://doi.org/10.1093/scipol/scy023
- Dickinson, J.L., Shirk, J., Bonter, D., Bonney, R., Crain, R.L., Martin, J., Phillips, T., Purcell, K., 2012. The current state of citizen science as a tool for ecological research and public engagement. Front. Ecol. Environ. 10, 291–297. https://doi.org/10.1890/110236
- Douthwaite, B., Alvarez, S., Cook, S., Davies, R., George, P., Howell, J., Mackay, R., Rubiano, J., 2007. Participatory Impact Pathways Analysis: A Practical Application of Program Theory in Research-for-Development. Can. J. Program Eval. Tor. 22.
- Dunn, M.B., Jones, C., 2010. Institutional logics and institutional pluralism: The contestation of care and science logics in medical education, 1967–2005. Adm. Sci. Q. 55, 114–149.
- Ely, A., Van Zwanenberg, P., Stirling, A., 2014. Broadening out and opening up technology assessment: Approaches to enhance international development, co-ordination and democratisation. Res. Policy 43, 505–518. https://doi.org/10.1016/j.respol.2013.09.004
- Etzkowitz, H., Leydesdorff, L., 1998. The endless transition: A "triple helix" of university-industry-government relations. Minerva 36, 203–208. Franzoni, C., Scellato, G., Stephan, P., 2014. The mover's advantage: The superior performance of migrant scientists. Econ. Lett. 122, 89–93. https://doi.org/10.1016/j.econlet.2013.10.040
- Franzoni, C., Scellato, G., Stephan, P., 2012. Foreign-born scientists: mobility patterns for 16 countries. Nat. Biotechnol. 30, 1250–1253. Freeman, R.B., 2003. Does globalization of the scientific/engineering workforce threathen U.S. economic leadership? NBER Work. Pap. Ser.
- Freeman, R.B., Huang, W., 2014a. Collaborating With People Like Me: Ethnic co-authorship within the US. National Bureau of Economic Research.
- Freeman, R.B., Huang, W., 2014b. Collaboration: Strength in diversity. Nature 513, 305–305. https://doi.org/10.1038/513305a

- Frodeman, R., Parker, J., 2009. Intellectual merit and broader impact: The National Science Foundation's broader impacts criterion and the question of peer review. Soc. Epistemol. 23, 337–345.
- Gersick, C.J., Dutton, J.E., Bartunek, J.M., 2000. Learning from academia: The importance of relationships in professional life. Acad. Manage. J. 43, 1026–1044. Gläser, J., Laudel, G., 2007. The Social Construction Of Bibliometric Evaluations, in: Whitley, R.,
- Gläser, J. (Eds.), The Changing Governance of the Sciences, Sociology of the Sciences Yearbook. Springer Netherlands, pp. 101–123. https://doi.org/10.1007/978-1-4020-6746-4_5
- Hayashi, T., 2003. Effect of R&D programmes on the formation of university–industry–government networks: comparative analysis of Japanese R&D programmes. Res. Policy 32, 1421–1442. https://doi.org/10.1016/S0048- 7333(02)00158-0
- Holbrook, J.B., 2012. Re-assessing the science-society relation: The case of the US National Science Foundation's broader impacts merit review criterion (1997-2011). Techonology Soc. 27, 437–451.
- Jonkers, K., Cruz-Castro, L., 2013. Research upon return: The effect of international mobility on scientific ties, production and impact. Res. Policy 42, 1366–1377. https://doi.org/10.1016/j.respol.2013.05.005
- Katz, J.S., Martin, B.R., 1997. What is research collaboration? Res. Policy 26, 1–18. https://doi.org/10.1016/S0048- 7333(96)00917-1 Kerr, S.P.,
- Kerr, W., Özden, Ç., Parsons, C., 2016. Global Talent Flows. J. Econ. Perspect. 30, 83–106. https://doi.org/10.1257/jep.30.4.83
- Khadria, B., 2002. Shifting Paradigms of Globalization: The Twenty-first Century Transition Towards Generics in Skilled Migration from India. Int. Migr. 39, 45–71. https://doi.org/10.1111/1468-2435.00171
- Kim, D., Twombly, S., Wolf-Wendel, L., 2012. International Faculty in American Universities: Experiences of Academic Life, Productivity, and Career Mobility. New Dir. Institutional Res. 155, 27–46. https://doi.org/10.1002/ir
- Krackhardt, D., Stern, R.N., 1988. Informal networks and organizational crises: An experimental simulation. Soc. Psychol. Q. 123–140.
- Lamanna, F., Lenormand, M., Salas-Olmedo, M.H., Romanillos, G., Gonçalves, B., Ramasco, J.J., 2018. Immigrant community integration in world cities. PLOS ONE 13, e0191612. https://doi.org/10.1371/journal.pone.0191612

- Lebeau, L.-M., Laframboise, M.-C., Larivière, V., Gingras, Y., 2008. The effect of university–industry collaboration on the scientific impact of publications: the Canadian case, 1980–2005. Res. Eval. 17, 227–232. https://doi.org/10.3152/095820208X331685
- Lee, M.-D.P., Lounsbury, M., 2015. Filtering Institutional Logics: Community Logic Variation and Differential Responses to the Institutional Complexity of Toxic Waste. Organ. Sci. 26, 847–866. https://doi.org/10.1287/orsc.2014.0959
- Leydesdorff, L., Etzkowitz, H., 2003. Can 'the public'be considered as a fourth helix in university-industrygovernment relations? Report on the Fourth Triple Helix Conference, 2002. Sci. Public Policy 30, 55–61.
- Libaers, D., 2014. Foreign-Born Academic Scientists and Their Interactions with Industry: Implications for University Technology Commercialization and Corporate Innovation Management. J. Prod. Innov. Manag. 31, 346–360. https://doi.org/10.1111/jpim.12099
- Light, I., Sabagh, G., Bozorgmehr, M., Der-Martirosian, C., 1995. Ethnic economy or ethnic enclave economy. New Migr. Marketpl. Boston's Ethn. Entrep. 23.
- Lin, Z., Pearce, R., Wang, W., 2009. Imported talents: Demographic characteristics, achievement and job satisfaction of foreign born full time faculty in four-year American colleges. High. Educ. 57, 703–721. https://doi.org/10.1007/s10734-008-9171-z
- Link, A.N., Siegel, D.S., Bozeman, B., 2007. An empirical analysis of the propensity of academics to engage in informal university technology transfer. Ind. Corp. Change 16, 641–655. https://doi.org/10.1093/icc/dtm020
- Llopis, O., D'Este, P., 2016. Beneficiary contact and innovation: The relation between contact with patients and medical innovation under different institutional logics. Res. Policy 45, 1512–1523. https://doi.org/10.1016/j.respol.2016.03.004
- Mamiseishvili, K., Rosser, V.J., 2010. International and citizen faculty in the United States: An examination of their productivity at research universities. Res. High. Educ. 51, 88–107. https://doi.org/10.1007/s11162-009-9145-8
- McDonald, R., Cheraghi-Sohi, S., Bayes, S., Morriss, R., Kai, J., 2013. Competing and coexisting logics in the changing field of English general medical practice. Soc. Sci. Med. 93, 47–54. https://doi.org/10.1016/j.socscimed.2013.06.010
- Miller, C. A., & Wyborn, C. (2020). Co-production in global sustainability: Histories and theories. Environmental Science & Policy, 113, 88–95. https://doi.org/10.1016/j.envsci.2018.01.016
- McPherson, M., Smith-Lovin, L., Cook, J.M., 2001. Birds of a Feather: Homophily in Social Networks. Annu. Rev. Sociol. 27, 415–444. https://doi.org/10.1146/annurev.soc.27.1.415

- Melkers, J., Kiopa, A., 2010. The Social Capital of Global Ties in Science: The Added Value of International Collaboration. Rev. Policy Res. 27, 389–414. https://doi.org/10.1111/j.1541-1338.2010.00448.x
- Merton, R.K., 1973. The sociology of science: Theoretical and empirical investigations. University of Chicago press.
- NAFSA: Association of International Educators, 2006. Restoring U.S. competitiveness for international students and scholars. NAFSA: Association of International Educators.
- Narin, F., 1976. Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity. Computer Horizons Washington, D. C. N
- erad, M., 2010. Globalization and the Internationalization of Graduate Education: A Macro and Micro View. Can. J. High. Educ. 40, 1–12.
- OECD, 2008. The global competition for talent: Mobility of the highly skilled. OECD, Paris.
- Owen-Smith, J., 2003. From separate systems to a hybrid order: accumulative advantage across public and private science at Research One universities. Res. Policy 32, 1081–1104. https://doi.org/10.1016/S0048-7333(02)00111-7
- Perkmann, M., Salter, A., Tartari, V., 2011. Reaching across institutional logics: Arbitrage vs. contamination. DRUID 2011 Innov. Strategy Struct. Inst. Syst. Reg. Cph. Bus. Sch. 15–17.
- Perkmann, M., Tartari, V., McKelvey, M., Autio, E., Broström, A., D'Este, P., Fini, R., Geuna, A., Grimaldi, R., Hughes, A., Krabel, S., Kitson, M., Llerena, P., Lissoni, F., Salter, A., Sobrero, M., 2013. Academic engagement and commercialisation: A review of the literature on university–industry relations. Res. Policy 42, 423–442. https://doi.org/10.1016/j.respol.2012.09.007
- Perkmann, M., Walsh, K., 2008. Engaging the scholar: Three types of academic consulting and their impact on universities and industry. Res. Policy, Special Section Knowledge Dynamics out of Balance: Knowledge Biased, Skewed and Unmatched 37, 1884–1891. https://doi.org/10.1016/j.respol.2008.07.009
- Polanyi, M., 2000. The republic of science: Its political and economic theory. Minerva 38, 1–21.
- Rafols, I., Molas-Gallart, J., Woolley, R., 2015. Science and Technology Indicators In & For the Peripheries. A Research Agenda. Presented at the Proceedings of ISSI 2015 Istanbul: 15th International Society of Scientometrics and Informetrics Conference, Istanbul, Turkey, pp. 560–564.

- Rahm, D., Bozeman, B., Crow, M., 1988. Domestic Technology Transfer and Competitiveness: An Empirical Assessment of Roles of University and Governmental R&D Laboratories. Public Adm. Rev. 48, 969–978. https://doi.org/10.2307/976993
- Ramos-Vielba, I., Sánchez-Barrioluengo, M., Woolley, R., 2016. Scientific research groups' cooperation with firms and government agencies: motivations and barriers. J. Technol. Transf. 41, 558–585. https://doi.org/10.1007/s10961- 015-9429-4
- Rogers, J.D., Bozeman, B., 2001. "Knowledge Value Alliances": An Alternative to the R&D Project Focus in Evaluation. Sci. Technol. Hum. Values 26, 23–55.
- Ryan, L., Sales, R., Tilki, M., Siara, B., 2008. Social networks, social support and social capital: The experiences of recent Polish migrants in London. Sociology 42, 672–690.
- Sabharwal, M., Varma, R., 2015. Scientific Diaspora: Stay Plans of Indian Faculty in the United States. Perspect. Glob. Dev. Technol. 14, 368–387. https://doi.org/10.1163/15691497-12341351
- Sarewitz, D., 2016. Saving Science. New Atlantis 49, 4–40. Sauermann, H., Stephan, P., 2012. Conflicting Logics? A Multidimensional View of Industrial and Academic Science. Organ. Sci. 24, 889–909. https://doi.org/10.1287/orsc.1120.0769
- Scott, P., 2015. Dynamics of Academic Mobility: Hegemonic Internationalisation or Fluid Globalisation. Eur. Rev. 23, S55–S69. https://doi.org/10.1017/S1062798714000775
- Séguin, B., Singer, P.A., Daar, A.S., 2006. Scientific diasporas. Sci.-N. Y. THEN Wash.-5780, 1602.
- Spaapen, J., Drooge, L. van, 2011. Introducing 'productive interactions' in social impact assessment. Res. Eval. 20, 211–218. https://doi.org/10.3152/095820211X12941371876742
- Stephan, P., Franzoni, C., Scellato, G., 2016. Global competition for scientific talent: evidence from location decisions of PhDs and postdocs in 16 countries. Ind. Corp. Change 25, 457–485. https://doi.org/10.1093/icc/dtv037
- Sugimoto, C.R., Robinson-Garcia, N., Murray, D.S., Yegros-Yegros, A., Costas, R., Larivière, V., 2017. Scientists have most impact when they're free to move. Nature 550, 29. https://doi.org/10.1038/550029a
- Teichler, U., 2015. Academic Mobility and Migration: What We Know and What We Do Not Know. Eur. Rev. 23, S6–S37. https://doi.org/10.1017/S1062798714000787
- Thornton, P.H., Ocasio, W., Lounsbury, M., 2015. The Institutional Logics Perspective, in: Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource. John Wiley & Sons, Inc. https://doi.org/10.1002/9781118900772.etrds0187

- Tijssen, R.J.W., Yegros-Yegros, A., Winnink, J.J., 2016. University–industry R&D linkage metrics: validity and applicability in world university rankings. Scientometrics 109, 677–696. https://doi.org/10.1007/s11192-016-2098-8 T
- Trencher, G., Yarime, M., McCormick, K.B., Doll, C.N.H., Kraines, S.B., 2014. Beyond the third mission: Exploring the emerging university function of co-creation for sustainability. Sci. Public Policy 41, 151–179. https://doi.org/10.1093/scipol/sct044
- van Leeuwen, T.N., 2005. Descriptive versus evaluative bibliometrics, in: Handbook of Quantitative Science and Technology Research. Springer, pp. 373–388.
- Varma, R., 2010. India-Born in the U.S. Science and Engineering Workforce. Am. Behav. Sci. 53, 1064–1078. https://doi.org/10.1177/0002764209356239
- Wagner, C.S., Jonkers, K., 2017. Open countries have strong science. Nature 550, 32. https://doi.org/10.1038/550032a
- Wagner, C.S., Leydesdorff, L., 2005. Network structure, self-organization, and the growth of international collaboration in science. Res. Policy 34, 1608–1618. https://doi.org/10.1016/j.respol.2005.08.002
- Wang, J., Berzins, K., Hicks, D., Melkers, J., Xiao, F., Pinheiro, D., 2012. A boosted-trees method for name disambiguation. Scientometrics 93, 391–411. https://doi.org/10.1007/s11192-012-0681-1
- Whitmer, A., Ogden, L., Lawton, J., Sturner, P., Groffman, P.M., Schneider, L., Hart, D., Halpern, B., Schlesinger, W., Raciti, S., 2010. The engaged university: providing a platform for research that transforms society. Front. Ecol. Environ. 8, 314–321.

Redefining Success and Quality in Basic and Applied Science

Emanuela Reale, Andrea Vargiu, Andrea Orazio Spinello, Valentina Ghibellini
CNR-IRCRES, Italy
University of Sassari, Italy
CNR-IRCRES, Italy
University of Sassari, Italy

ASSESSING TRANSFORMATIVE COMMUNITY ENGAGEMENT. OPEN ISSUES ANDCHALLENGES

ABSTRACT

One of the key activities within the university's third mission is public engagement (PE). Despite its relevance in almost all the European universities, the notion of PE is still under examination because the types of activities that can be labelled as such are extremely different as to values, actors involved, approaches, and fields (Miller, 2001; NCCPE, 2010; Grand et al., 2015; Watermeyer and Lewis, 2018).

In this paper, we address some crucial open issues related to the evaluation of a specific form of engagement: community engagement (CE) under a comparative perspective. We will notably focus on assessing its capability as to being transformative of society and of university organization. In order to do so, we will build upon the theoretical scaffolding of aPRIN2022 project title PLAACES(Portraits and Landscapes of Academic Community Engaged Scholarship).

We define CE as a form of engagement that is characterised by reciprocity (regulatory principle), an emancipative prevailing interest, and a prevailing critical epistemology. Within the wider spectrum of PE, CE can have important overlap with the state engagement when research contributing to evidence-based policy is concerned. Here the most important element is the type of participation, which might imply involvement and co-production in both community and state engagement (Vargiu, 2014).

Literature generally defines being transformative as the capacity to bring about a durable, significant, and disruptive change. This approach is paired by a different one where being transformative is the capability to empower social actors for producing a lasting change in society and the capability of the involved university to change institutional settings and strategic agency. Empowering social actors means to develop mutual learning and actions going beyond sharing information and understanding. Thus, underthe former conceptualization there is a clear overlapping with impact; on the contrary,following to the latterconceptualization transformation cannot be assimilated to social impact. (Stilgoe et al., 2014; Bucchi, 2008; Johnson, 2020; Reale, 2022; Murunga, 2022).

There are two further important elements to consider when evaluation is concerned. One is the degree of institutionalisation of CE, since ‘engagement is an emergent outcome that

imust continually be reaffirmed in its institutional settings'. (Benneworth et al., 2009). The second is the barriers that universities can face, which derive from excluded communities, inequalities, bureaucratic fulfilments, funding patterns, casualization of the research work, and academic rewards. These elements can be serious constraints to community engagement. (Vargiu, 2014; Benneworth et al., 2013; Ruiz Bravo, 1992; Carney and Oliver, 2018; Heney and Poleykett, 2021).

Measuring and assessing the effectiveness of the public engagement on society is therefore a great challenge; currently the evaluation practice is mainly shaped by the government evaluation agencies within the massive national evaluation exercises of the universities. In this paper, we want to put in context the problems with this new frontier for university evaluation and present first insights on existing differences at the national level and in the national systems of higher education that can affect the possibility of community engagement to be transformative toward academic institutions and society.

The method used is a case study approach, based in secondary data,documentation,andindicatorstoshaping the different configurations ofsix universities withinthree European countries, two from thecontinentalEurope (France and Italy) and one from the Ang Saxon tradition (UK).Case studies also foresee a set of interviews at institutional levels and life stories at individual levels to capture the practices of public engagement and its transformative effects.

The study is under development. Preliminary findings will present the different gaps in the literature dealing with community transformative engagement and how the study tries to address them. Furthermore, data on the different national policies towardHEIs engagement are collected, as well as data on the different university strategies toward community engagement.

The analysis of the materials is aimed at pointing out and discussing factors at national and system levels that, according to the literature (Benneworth and Jongbloed 2013), can promote or constrain social community engagement. A special attention is devoted to national planning and institutional strategies promoting social engagement, social engagement as a core element in the governance of the HEI, financial incentives (dedicated government funding streams, core funding allocation, special rules for attracting students), skills for engagement (rewarding of staff by HEIs for community engagement in terms of career development and promotion, participation and co-creation of knowledge), and measures to promote the regional embeddedness of the universities (linkages with the social communities and the economic actors of the region where the university is located).

All these elements can influence the evaluation of engagement and the likelihood that engagement will be transformative for society and academic organizations. Evaluating the drivers and conditions for transformative community engagement is a means of changing the nature of the university mission so that being transformative can become the overall goal of academia. As a matter of fact, the issue of community transformative engagement in universities is still largely unexplored. The added value of the paper rests on the capability to supply evidence on how evaluation can facilitate or hinder the capability of the universities to commit themselves to community engagement and how to produce transformative effects on universities and on society.

Florian Wittmann, Sarah Seus, Valentin Wagner
Fraunhofer ISI, Germany
Fraunhofer ISI, Germany
Institut für Innovation und Technik, Germany

DEVELOPMENT OF AN IMPACT-ORIENTED MONITORING SYSTEM FOR A RESEARCHFUNDING ORGANISATION

ABSTRACT

Along with the increased attention to the societal impact of science, also organisations funding research activities, including research fellowship providers, have been increasingly subjected to the demand of better understanding the impacts of their funding activities (Bornmann 2013; Sørensen et al. 2022). At the same time, the idea of impact pathways as basic descriptors of different types of research and different logics and activities in research has gained popularity in different contexts (Bruno et al. 2019; Griniece et al. 2020; Muhonen et al. 2019).

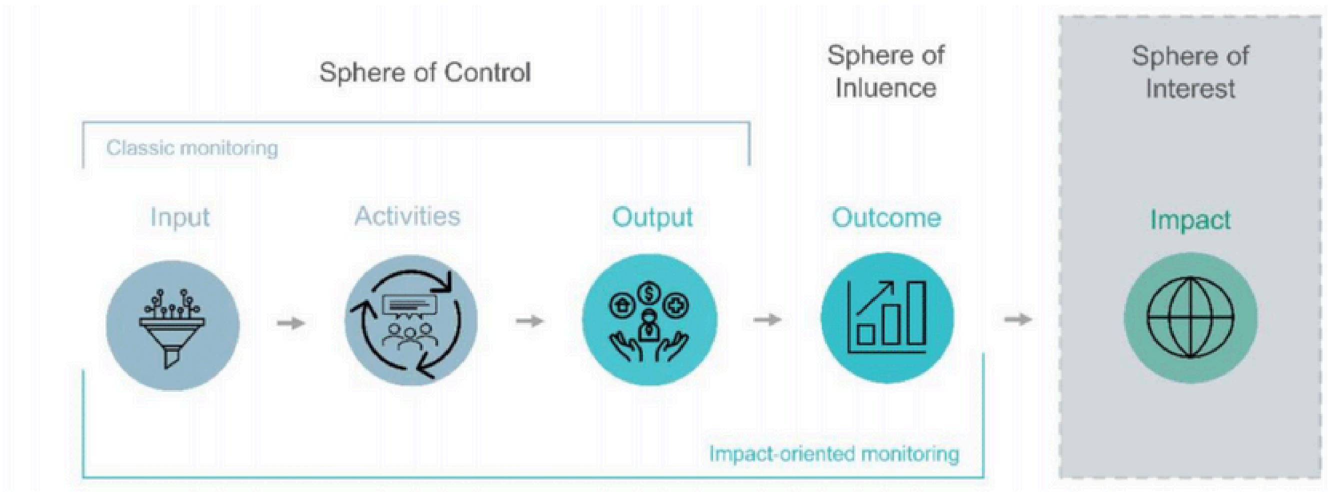
This contribution aims to present the approach for developing a consistent framework for impact-orientated monitoring at different organisational levels (overarching funding portfolio/activities, individual funding schemes) for research funding ganizations. The novelty is its modular structure by making use of the idea of generic impact pathways as the key building blocks. These impact pathways describe fundamental logics of funding activities, targeting different key stakeholder groups (individual researchers, universities, etc.) that can be flexibly combined and adjusted to different funding schemes and programs. Thereby, this contribution builds on the experiences of the Fraunhofer Institute for Systems and Innovation Research ISI in Karlsruhe and the Institut für Innovation und Technik-It that were gained through accompanying the Alexander von Humboldt Foundation in their efforts for setting up a framework for impact-orientated monitoring.

Alexander von Humboldt Foundation The 1953-established Alexander von Humboldt Foundation (AvH) is one of the key German organisations promoting international cultural dialogue and academic exchange. The AvH provides research fellowships/prizes (mainly) to researchers based in research organisations outside Germany. Among the key characteristics of the Alexander von Humboldt Foundation are: i) a large variety of long-standing and well-established funding programs targeting researchers at different career levels ii) a mission to strengthen collaborative academic relations and improve the internationalisation of the research location in Germany through incoming and outgoing programs iii) a strong network of more than 30,000 alumni worldwide,

We build on (Belcher et al. 2020) concept of influencing spheres expressing decreasing ability to control funding effects by donors and apply it to the entire funding portfolio of the Alexander von Humboldt Foundation. In combination with the approach to impact-

orientated monitoring (Griniece et al. 2019), we can define and delineate the monitoring framework. (cf. Figure 1). The idea of influencing spheres allows to define the parameters that the research funding organisation can control and, in consequence, influence: Inputs, activities, and outputs are grouped in the sphere of control, as the research funding body can heavily influence these three effect categories, e.g., through the design of the funding program or the selection of the projects or candidates. Second, a sphere of influence, i.e., outcomes that are still strongly tied to the funding impulse but are beyond the immediate control of the fellowship provider. Finally, the sphere of interest that comprises the anticipated impacts of funding activities that are beyond control (and scope of the monitoring system). In contrast to "classical" monitoring approaches, impact-orientated monitoring aims to include outcome-level indicators, thus going beyond the output level for indicators only.

FIGURE 1: MAIN LOGIC OF THE IOOI MODEL



Source: Own elaboration based on Belcher et al. (2020) and Griniece et al. (2020:19).

Based on strategic documents, program descriptions, and interviews with program managers and key representatives within the Alexander von Humboldt Foundation, we identify six fundamental objectives that can be developed into six key impact pathways to which the foundation's entire funding activities contribute. These pathways are:

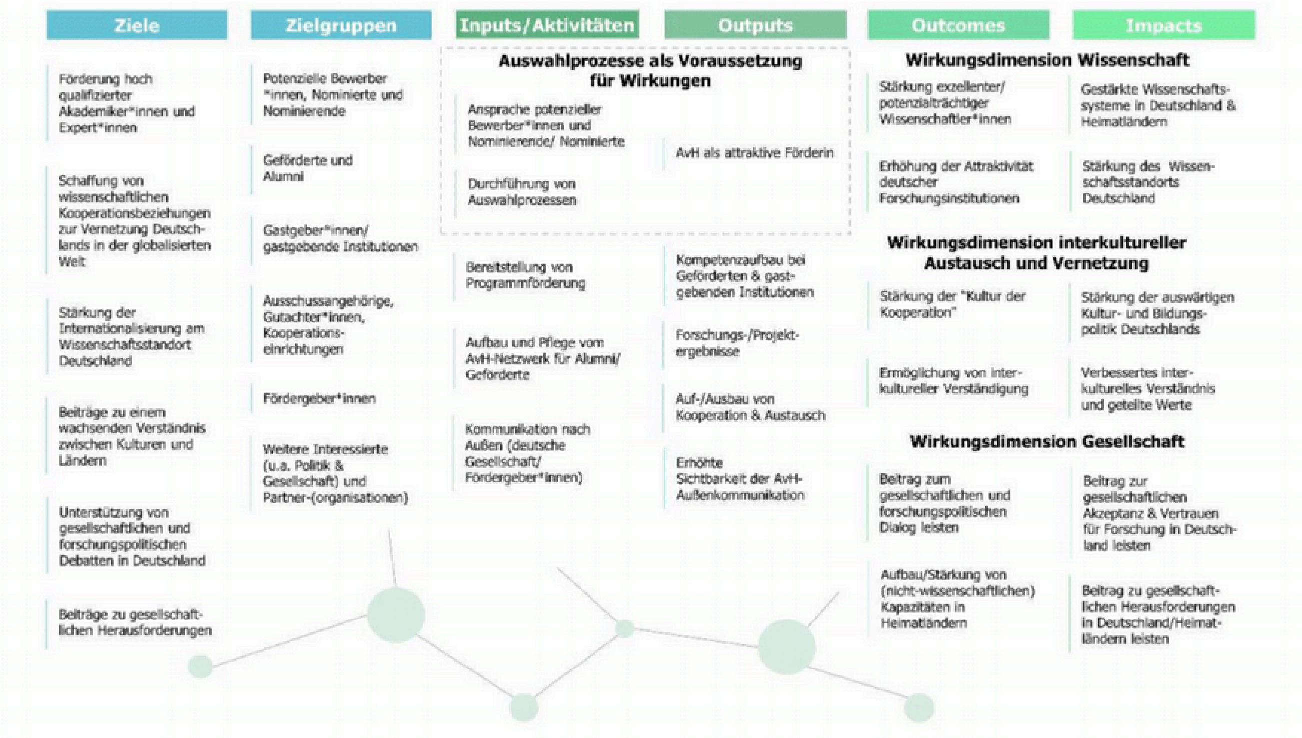
1) Strengthening excellent scientists (and young scientists with excellence potential): Through scholarships, individual researchers get the possibility to pursue a research project that strengthens their scientific qualifications, supports their career development, and thereby contributes to scientific progress.

2) Strengthening Germany as a research location: By providing scholarships, foreign researchers are invited to do research in a German research-performing organisation for

several years. This benefits the science system in Germany by increasing its internationalisation and visibility and facilitating potentially long-lasting relationships with international excellent researchers.
3) Strengthening a culture of cooperation: The uniqueness of the Humboldt foundations is the Humboldt network: a network of all current and former funded researchers. Activities such as networking events and former “Humboldtianer” functioning as hosts for new scholarship holders contribute to the network expansion and to a positive image of Germany as a research-promoting country.
4) Strengthening intercultural communication/understanding: At an individual level of the grant holders, their research stay abroad is expected to contribute to intercultural learning and experience but also to act as intermediaries between different countries and cultures.
5) Contribution to societal and research policy debates in Germany: The focus of this strand is particularly on steering societal debates (on science) in Germany and how the Alexander von Humboldt Foundation can contribute to ongoing debates in Germany.
6) Contributions to societal challenges Finally, researchers, with their research projects, can contribute in multiple ways to solving societal challenges.

This can range from relevant research in areas like sustainability to capacity building in less developed countries, Whereas impact pathways 1 and 2 refer to scientific impacts, impact pathways 3 and 4 describe impacts related to intercultural communication and exchange/cooperation. Finally, impact pathways 5 and 6 focus on the societal contributions of the Alexander von Humboldt Foundation. At the same time, these different impacts target specific groups of actors. Whereas impact pathways 1), 3), 4) and 6) are closely linked to individual researchers, impact pathways 2) takes a closer look at hosting universities in Germany, while impact pathway 5) considers the (science) communication activities of the Alexander von Humboldt Foundation. These impact pathways in combination can be subsumed in the main logic model of the funding activities of the Humboldt Foundation (cf. Figure 2)

FIGURE 2: MAIN LOGIC MODEL FOR THE ALEXANDER VON HUMBOLDT FOUNDATION:



We used the impact pathway approach to develop the overall intended effects and used it to build a framework for impact-orientated monitoring, including the development of suitable indicators. This approach offers multiple advantages:

First, it allows for a flexible modular approach, describing both key logics of funding activities at the level of individual scholarship/award programs as well as the level of the activities of the Alexander von Humboldt Foundation. With a few exceptions, all programs, despite different priorities, target groups, etc., can be described comprehensively by different combinations of these six impact pathways.

Second, the approach supports consistency across different projects by emphasising the underlying logic of anticipated impacts instead of idiosyncratic program characteristics, ensuring a similar reference point and the possibility for identification of potential blind spots in individual programs, as these traditionally favoured certain areas of impacts (e.g., scientific excellence), while e.g., paying less attention to the raising debates, such as societal impacts of research (funding). Thereby, the framework can also be utilised for the IOOI schemes for newly defined programs.

Third, building on the idea of similar logics across different programs despite differences in target groups, etc., allows to derive a limited set of indicators for each of these impact pathways that can be, with minor adjustments, used across the different programs of a funding portfolio. The use of comparable indicators, moreover, allows the aggregation of program-level data for an overarching monitoring going beyond individual programs.

While tailoring these impact pathways to the specific requirements of the Alexander von Humboldt Foundation, we consider these pathways as potentially relevant for other fellowship providers/organisations supporting international scientific exchange and research. In the context of this contribution, we therefore want to discuss the broader generalisability of these impact pathways. Moreover, a second aspect for discussion should be potential conceptual challenges and the needs for internal management/steering, providing different perspectives on selection processes.

REFERENCES:

- Belcher, B. M.; Claus, R.; Davel, R. (2020): A refined method for theory-based evaluation of the societalimpacts of research. In: MethodsX, 23 (7), pp. 1–20.
- Bornmann, L. (2013): What is societal impact of research and how can it be assess? A literature survey. In:Journal of the American Society for Information Science and Technology, 64 (2), pp. 217–233.
- Bruno, N.; Kadunc, M. (2019): Impact Pathways: Tracking and communicating the impact of the EuropeanFramework Programme for research and innovation. In: fteval Journal for Research and TechnologyPolicy Evaluation, 47, pp. 62–71.
- Griniece, E.; Angelis, J.; Reid, A. R.; Vignetti, S.; Catalano, J.; Helman, A.; Barberis Rami, M.; Kroll, H.(2020): Guidebook for Socio-Economic Impact Assessment of Research Infrastructures. Available athttps://doi.org/10.5281/zenodo.3950043.
- Muhonen, R.; Benneworth, P.; Olmos-Peñuela, J. (2019): From productive interactions to impactpathways: Understanding the key dimensions in developing SSH research societal impact. In: ResearchEvaluation.https://doi.org/10.1093/reseval/rvz003.
- Sørensen, O. H.; Bjørner, J.; Holtermann, A.; Dyreborg, J.; Sørli, J. B.; Kristiansen, J.; Nielsen, S. B. (2022):Measuring societal impact of research—Developing and validating an impact instrument foroccupational health and safety. In: Research Evaluation, 31 (1), pp. 118–131.https://doi.org/10.1093/reseval/rvab036.

Sarah Seus, Florian Wittmann, Nele Weiblen
Fraunhofer ISI, Germany

EXPLORING UNINTENDED CONSEQUENCES IN STI EVALUATIONS AND MONITORING

ABSTRACT

When conducting program evaluations, the focus is first of all on the intended effects that should be achieved by a funding impulse. Especially with the turn to mission orientation and transformative policies (Mazzucato 2018; Diercks et al. 2019), STI funding organisations have aligned to policy demands and have designed programs aiming at impacts far beyond the research sector (e.g., on the European level, the "Green Deal Calls" in H2020, in Germany, the Strategy for Research for Sustainability and the High-Tech Strategy 2025; the Swiss National Research Programmes (NFP) or the Swedish Strategic Innovation Programme (SIP). In the last years, the STI evaluation community has responded to the challenges that came with these new funding programs: an intensive discussion arose on the nature of intended effects of research funding, and new concepts and methods to model and measure these effects have been developed (Bruno und Kadunc 2019; Bührer et al. 2022; Spaapen et al. 2011; Dinges et al. 2020; Seus und Bührer 2021).

So far, only a few publications have conceptualised the issues of unintended consequences for the STIpolicy field. This might be a consequence of the fact that the concept of unintended effects remains hard to grasp: In some publications, these are only associated with negative effects, whereas other studies point to the existence of potential unintended positive effects (Jabeen 2018; Lorenc et al. 2019). However, there is a growing consensus that "the effect of an intervention other than those it aimed to achieve (Jabeen 2016, S. 144) matters. In evaluation studies, unintended effects are sometimes described, but these studies remain rare (and there is no common approach to how to assess them).

Our contributions therefore seek to explore and consolidate the so far dispersed contributions on unintended consequences by bringing them together in a review and exploring what type of unintended consequences are relevant for either monitoring and/or evaluation activities.

First, our contribution aims at giving an overview of the available literature tackling unintended consequences. We draw on literature that discusses and classifies unintended consequences of policy interventions in general and combine it with more specific STI literature on unintended consequences in research and research funding. We make use of a systematic literature research approach combined with the snowball principle and analysis of an evaluation database (SIPER) to collect both conceptual as well as empirical contributions focusing on unintended effects of STI activities.

Based on the literature review, we develop a classification of unintended consequences,

depending on different characteristics that unintended consequences show. This classification first allows to better understand the diversity of unintended consequences and their distinct specificities. Secondly, we are taking this classification as a starting point for reflecting about the relevance of different types of unintended consequences for monitoring and evaluation.
We plan to validate our classification empirically using end-of-project reporting of research projects funded by the Volkswagen Foundation, a German research funding organisation.

Within the literature, unintended consequences have been conceptualized according to differentcriteria and dimensions.
A first dimension concerns the knowability of effects. Unintended effects, according to this dimension, can be anticipated, "known by the actor at the time of action (Jabeen 2018, S. 264), or unanticipated. In the case of unanticipated consequences, a further division between foreseeable and unforeseeable effects is made. This division related to the idea that effects can either be predicted or not (Jabeen 2018). Unforeseeable effects occur where "adaptive and nonlinear phenomena make prognostication impossible (Morell 2005, S. 445; Braun 2009). In contrast, foreseeable or predictable effects might still not be foreseen, leading to unforeseen effects, especially in those cases for which applicable analytical frameworks and experience were not considered (Morell 2005, S. 446). A last subcategory of effects in the knowability dimension, although less discussed in the literature, are overlooked effects, meaning effects that are "known but deliberately ignored for practical, political, or ideological reasons (Morell 2005, S. 445).
A second literature strand discusses unintended consequences with regards to the size of the unintended effects. Meijer und Sivertsen 2020) discuss the societal impact of research that can either be normal or extraordinary. Normalsocietal impacts refer to "the results of active, productive, and responsible interactions between (units of) research organisations and other organisations according to their purposes and aims in society (Meijer und Sivertsen 2020, S. 67). In contrast, extraordinary societal impacts are defined as "rare incidences where [...] interactions between science and society have unexpected widespread positive or negative implications for society (Meijer und Sivertsen 2020, S. 67).
Thirdly, unintended effects can be differentiated according to their value: the impact can be evaluated as positive, negative, or neutral (Bonell et al. 2019; Derrick et al. 2018; Jabeen 2016; Meijer und Sivertsen 2020). In combining the dimensions of size and value, Derrick et al. (2018) provide the example of 'Grimpacts', which are impacts of extraordinary size but

with negative implications for society. However, one should be cautious about generally equating unintended effects with negative effects, as they could also be beneficial.

A last dimension concerns the controllability of the unintended effects. Controllability refers to the fact that even if unintended effects are anticipated, it might not be possible to avoid them (Braun 2009). Following Braun's line of argumentation, the controllability of unintended effects hinges on further different parameters. Whereas simple effects, as a result of individual actions, are controllable, more complex or intricate effects will be more difficult to control.

No matter the size, value, knowability, or controllability of unintended effects, once they do occur, one can additionally examine how they are distributed, especially with regards to the concerned stakeholder and/or the timing of their occurrence (Jabeen 2018).

The following table displays the characteristics of unintended consequences as discussed in theliterature and attempts a classification. Five dichotomous categories have been identified that characterise the unintended effects. Crosscutting to these five categories, we found two cross-cutting categories (stakeholder concerned and timing) that could apply to all effect categories.

TABLE1:TYPESOFUNINTENDEDEFFECTS–FIRST CLASSIFICATION

Categorisation of the effect	Characteristics	Literature
Knowability of the effect	Anticipated vs. unanticipated Foreseeable vs. unforseeable	Jabeen 2018 Braun 2009 Morell 2005
Size of the effect	Normal vs extraordinary	Meijer und Sivertsen 2020
Value of the effect	Positive vs. negative vs. neutral	Jabeen 2016; Bonell et al. 2019; Derrick et al. 2018; Meijer und Sivertsen 2020 Braun 2009
controllability	Controllable vs. uncontrollable	Braun 2009
Stakeholder concerned by the effect	Participants /beneficiaries / external stakeholder / funding body / all / ….	Jabeen 2018
Timing when the effect occurs	During funding period / after project ends / …	Jabeen 2018

This classification attempt is a starting point to reflect on what different characteristics of unintendedconsequences are useful in the context of monitoring and evaluationactivitiesand consider thepotential relations between these different effect types.

Building on the obtained results, we seek to discuss what kind of unintended consequences can/should be considered in monitoring and evaluation activities (we assume that other/different unintended consequences might be more important for monitoring than for evaluation) and which one might be rather of theoretic relevance/but not in the main focus of monitoring/evaluationactivities.For example, normal-size negative effects that have not been intended might be of relevance for monitoring activities, while evaluation might rather be interested in extraordinary (grimpact) effects that were not (ex-ante) foreseeable but affect the implementation of a programconsiderably.

We want to empirically test the outlinedcategorisationof unintended consequences by using datafrom a German research funding organisation, the VolkswagenStiftung(monitoring) and the SIPERdatabase (evaluation). We will analyse the final project reporting of VolkswagenStiftung funded.We will use qualitative and quantitative text analysis to test the categories outlined in this abstract to critically reflect the validity of the framework and discuss potential modifications.

The analysis will be conducted during summer 2024 and be ready to present for the REvaluationConference in December 2024.

REFERENCES:

- www.si-per.eu
- Bonell, Chris; Lorenc, Theo; Oliver, Kathryn; Tinkler, Jane (2019): Understanding the unintendedconsequences of public health policies: the views of policymakers and evaluators.In:BMC PublicHealth19, Artikel 1057. Online verfügbar unterhttps://doi.org/10.1186/s12889-019-7389-6.
- Braun, Dietmar (2009): Evaluation und unintendierte Effekte—eine theoretische Reflexion. In:Hildegard Matthies und Dagmar Simon (Hg.): Wissenschaf unter Beobachtung. Effekte und Defektevon Evaluationen. Wiesbaden: VS Verlagfür Sozialwissenschaften, S. 103–124.Online verfügbar unterhttps://doi.org/10.1007/978-3-531-90863-2_7.
- Bruno, Nelly; Kadunc, Martina (2019): Impact Pathways: Tracking and communicating the impact ofthe European Framework Programme for research and innovation. In:fteval Journal for Research andTechnology Policy Evaluation(47), S. 62–71. DOI: 10.22163/fteval.2019.330.
- Bührer, Susanne; Feidenheimer, Alexander; Walz, Rainer; Lindner, Ralf; Beckert, Bernd; Wallwaey,Elisa (2022): Concepts and methods tomeasure societal impacts–an overview. Karlsruhe(Fraunhofer ISI Discussion Papers Innovation Systems and Policy Analysis, 74)

- Derrick, G. E.; Faria, R.; Benneworth, P.; Budtz-Petersen, D.; Sivertsen, G. (2018): Towardscharacterising negative impact: Introducing Grimpact. In: Proceedings of the 23rd InternationalConference on Science and, S. 1199–1213. Online verfügbar unterhttp://hdl.handle.net/1887/65230.
- Dierck s, Gijs; Larsen, Henrik; Steward, Fred (2019): Transformative innovation policy: Addressingvariety in an emerging policy paradigm. In:Research Policy48 (4), S. 880–894. DOI:10.1016/j.respol.2018.10.028.
- Dinges, M.; Meyer, Susanne; Brodnik, Christoph (2020): Key Elements of Evaluation Frameworks forTransformative R&I Programmes in Europe. In:fteval Journal for Research and Technology PolicyEvaluation, S. 26–40. DOI: 10.22163/fteval.2020.486.
- Jabeen, Sumera (2016): Do we really care about unintended outcomes? An analysis of evaluationtheory and practice. In:Evaluation and Program Planning55, S. 144–154. Online verfügbar unterhttps://doi.org/10.1016/j.evalprogplan.2015.12.010.
- Jabeen, Sumera (2018): Unintended outcomes evaluation approach: A plausible way to evaluateunintended outcomes of social development programmes. In:Evaluation andProgram Planning68, S.262–274. Online verfügbar unterhttps://doi.org/10.1016/j.evalprogplan.2017.09.005
- Lorenc, Theo; Oliver, Kathryn; Tinkler, Jane (2019): Evaluating unintended consequences: Newinsights into solving practical, ethical and politicalchallenges of evaluation. In:Evaluation26, S. 61–75.Online verfügbar unterhttps://doi.org/10.1177/1356389019850847.
- Mazzucato, Mariana (2018): Mission-Oriented Mission-Oriented Research & Innovation in theEuropean Union. A problem-solving approach tofuel innovation-led growth. Hg. v. EuropeanCommission Directorate-General for Research, zuletzt geprüft am 21.05.2022.9
- Meijer, Ingeborg; Sivertsen, Gunnar (2020): Normal versus extraordinary societal impact: how tounderstand, evaluate, and improve research activities in their relations to society? In:ResearchEvaluation29 (1), S. 66–70. Online verfügbar unter 10.1093/reseval/rvz032.
- Morell, Jonathan A. (2005): Why Are There Unintended Consequences of Program Action, andWhat Are the Implications for Doing Evaluation? In:American Journal of Evaluation26 (4), S. 444–463.Online verfügbar unter 10.1177/1098214005281354.
- Seus, Sarah; Bührer, Susanne (2021): How to Evaluate a Transition-Oriented Funding Programme?Lessons Learned from the Evaluation of FONA, the German Framework Programme to PromoteSustainability Research. In:fteval Journal for Research and Technology Policy Evaluation(52), S. 10–18.DOI: 10.22163/fteval.2021.514.

- Spaapen, J.; van Drooge, L.; Propp, T.; van der Meulen, B.; Shinn, T.;Marcovich, A. (2011): Socialimpact assessment methods for research and funding instruments through the study of productiveinteractions (SIAMPI). Final report on social impacts of research. Hg. v. Royal Netherlands Academyof Arts and Sciences. Amsterdam

Cristina Rosemberg Montes, Neil Brown
Technopolis Group, United Kingdom

A HOLISTIC APPROACH TO MEASURING THE IMPACTS OF RESEARCH PROGRAMMES: THECASE OF THE UK STRATEGIC PRIORITIES FUND

ABSTRACT

The UK Strategic Priorities Fund (SPF) is an £831m (€1bn) fund from UK Research and Innovation (UKRI) that was announced in the Industrial Strategy White Paper in 2017. It was part of a wider package of measures, with SPF providing discovery-led R&I to complement challenge-led elements.

SPF was created to strengthen the UK's R&I capacity and address gaps in the funding system identified in the 2015 independent review by Nobel laureate Sir Paul Nurse 4. The review pointed out various issues related to the (alack of) awareness and coordination of strategic research efforts across research councils and government, gaps in the funding support for multi- and inter-disciplinary research (MIDRI), and the limited ability to respond quickly and materially to emerging challenges or opportunities. The review led to the creation of UKRI in 2018 (bringing together the seven Research Councils and the Innovation Agency) and a series of cross-council initiatives, including SFP.

SPF was an attempt to fund activities that tackle the main criticisms made in the review. Success was therefore defined by the ability of the Fund to effectively support multidisciplinary teams to address complex challenges, resulting in research outcomes that were relevant to government departments R&I needs. The scale of funding meant that SPF had the capacity to engage across the UK R&I system, and it has supported 34 programmes, encompassing a diverse range of R&I activities at various stages of maturity. This includes thematic programs from across the broad spectrum of R&I areas, from florology & biomedicine to AI & digital solutions and infrastructures, sustainability, and environment.

Our fund-level evaluation of SPF (led by Technopolis with Ipsos, CECAN, and Science Metrix) seeks to demonstrate what the Fund has delivered, build the evidence base on what works, and help improve this and other future similar initiatives. It has so far delivered an evaluation framework, baseline, and interim impact report (the subject of our presentation).

The Fund itself is novel:

- Challenge-driven: It includes a portfolio of 34 programs that encompass a diverse range of R&I activities at various stages of maturity. The common thread is the ambition to address complex challenges and some of the more pressing R&I questions of government departments.

- Multi-partner, multi-stakeholders approach: These programmes are implemented by lead organisations (UKRI Councils and/or PSREs) in coordination with other partners (UKRI Councils, Government Departments, and/or PSREs, plus other partners) through large collaborative projects, open competitions, investments in infrastructure, and support for the activities of institutes and centers.
- Additionally, the fund allows government departments Public Sector Research Establishments (PSREs) to submit SPF program proposals (as an unusual feature of SPF's design).
- Connecting with government R&I needs: SPF also helped establish a novel process for engagement between councils and government departments to identify, prioritise, and co-create relevant program ideas. This involved a multi-step prioritisation process to identify and then consolidate cross-departmental R&I priorities, resulting in a final list of priority proposals that was taken into account in the subsequent SPF selection process (with a 'prioritised bid' providing a strong rationale for selection by the panel).

The evaluation also offers a novel approach:

- It includes and implements a framework to evaluate the Fund portfolio level, covering a diversity of programs of £10m to £70m each, each covering multiple activities.
- It implements a mixed-methods approach to measure multidisciplinarity across the 'life-cycle of research, from application to team composition to integration of knowledge. Four different methods are employed to explore the extent of MIDRI in relation to SPF program composition, proposals, projects, project teams, and outputs, as well as the extent to which actions have been taken to support and enable MIDRI. Bibliometric analysis has been a key aspect of this approach, where we have looked at the diversity of the disciplinary backgrounds of SPF project teams and of co-authors on SPF papers, as well as at the diversity of wider knowledge that is being cited and mobilised in these SPF publications.
- It implements a counterfactual approach to test the extent to which the Fund is delivering in excess of what would have happened otherwise across its two key objectives (multidisciplinarity and uptake of research outcomes).

The evaluation included interviews with 160+ stakeholders (covering representatives from UKRI, academia, government departments and agencies, PSREs, charities, and industry.), 8 longitudinal case studies, combining more traditional methods (surveys, interviews) with more novel approaches (machine learning) and a data-driven approach to measure multi- and interdisciplinary at scale.

SupportingMIDRI: We found overall that SPF is helping to drive an increase in high-quality MIDRI at all stages, from applications and research teams through to publications and synthesis and dissemination activities. SPF has enabled this by making funding available for programs that intend to support and encourage MIDRI, as well as challenge-led programs that embed participation of multiple councils and different stakeholders across the program lifecycle.

We have seen widespread efforts across the portfolio to encourage MIDRI proposals, with this being explicitly encouraged in most competitive calls. In fact, MIDRI has been required within calls for nearly half of programs. There have also been other events and activities to encourage and support MIDRI applications, such as to bring different disciplines together. These efforts appear to have been successful, with high rates of MIDRI applications. However, it is worth noting that this is only slightly higher than the overall UKRI average in the same period, which would support the idea (expressed via interviews) that councils are active in encouraging MIDRI more generally (not just through SPF).

Following on from the applications, we also found that nearly all SPF programs have put in place processes specifically designed for the assessment of MIDRI, with more than half of programs suggesting that these processes were new or enhanced, compared to what had previously been used.

The study goes on to show how this has translated into a high degree of MIDRI among SPF projects and the research teams. Results for SPF papers are consistently much higher than for UK or UKRI papers in relation to a series of indicators for multidisciplinarity. When we look at the other publications of SPF authors (either prior to SPF or in parallel to this program), the MIDRI indicators are also above average (suggesting SPF is attracting a high proportion of researchers alread active inMIDRI research). However, the figures are still higher for the SPF papers (suggesting further developing the extent of multidisciplinarity—even amongst those already active).

Finally, there are also early indications that SPF's MIDRI focus could have longer-term and wider (ecosystem) effects, with improved confidence reported amongst researchers and innovators and amongst funders in relation to applying for and investing in MIDRI.

Our evidence suggests that SPF is helping to effectively link up with and address government R&I priorities through providing additional funding in these priority areas, encouraging government involvement in programs and projects, and through the dissemination activities of programs plus the uptake of outputs for use in policy and decision-making.

Moreover, SPF helped establish a novel process (in Wave 2) for engagement between councils and government departments to identify, prioritise, and co-create relevant program ideas. This involved a multi-step prioritisation process to identify and then consolidate cross-departmental R&I priorities, resulting in a final list of 18 priority proposals that was considered in the SPF selection process.

Our study also shows that SPF has increased UKRI spend in a number of government R&I priority areas, bringing additional resource and focus. There are 4 areas where SPF has made a substantial difference (in terms of average value of funding): bacterial plant diseases, space weather, and research productivity, where SPF explains 56%–75% of overall increases in the average annual value of grants. Additionally, there is a 23%–35% increase across the other four priorities driven by SPF funding.

The Fund has also encouraged programs to involve government departments and agencies, both in the design and scoping of these programs but also in their ongoing implementation and governance. Most programs (27/33, or 82%) had some degree of government involvement in preparatory stages, with nearly half (48%) reporting strong government involvement. In consulting with program leaders again a year later (2022), we found that this breadth and depth of government involvement had been maintained, with 83% of programs reporting some involvement during program implementation and 45% reporting strong involvement. This is despite challenges that were also noted by many of the program leads relating to changes to personnel over time (in government departments and within partner organisations) and the competing demands on the time of government representatives (where programs were often engaging at quite a senior level at the bid stage).

The SPF experience and evaluation evidence provide useful learning with regards to supporting and enabling MIDRI or addressing government R&I priorities.

Future interventions aimed at supporting high-quality MIDRI might consider:

- The need to further standardise processes across councils to facilitate funder collaboration, as well as improve access to opportunities, regardless of the community and leading council involved.
- Allowing sufficient time for 'MIDRI partnerships' to form and to respond to calls—particularly if there is a desire to bring together communities that do not usually collaborate.
- Providing funding for MIDRI networks to form and to undertake preliminary activities, bringingtogether different communities, and exploring new ideas for future projects.
- Allowing more time for the grants, for additional learning and interaction across disciplines

There are also various lessons that have been learnt within programs around the challenges of setting up a MIDRI-appropriate assessment process (in terms of the reviewers involved, or the criteria and guidance to assess MIDRI), which could usefully be shared and socialised across councils.

- Future interventions aiming at addressing government R&I priorities may also consider:
- Building upon the process of identifying government priorities as part of the second wave of program selection, which is generally felt to have been successful and should be repeated.
- Planning for and dedicating resources through the life cycle of the program to maintain engagement with relevant policy makers and other end users (to align research and ensure its quality and relevance). In particular, we have found evidence that maintaining user engagement throughout a program can be a challenge. Consequently, conscious efforts are needed to ensure re-engagement in later phases,

including a special focus on the nature, style, and focus of outputs and dissemination activities (e.g., developing synthesised policy briefs rather than (or in addition to) individual academic publications).

Dr Elizabeth Gadd, Laura Himanen, Erica Conte, Marianne Gauffriau, Tanja Strom, Baron Wolf
INORMS ResearchEvaluation Group / Loughborough University, United Kingdom,
CSC–IT Center for Science, Finland
Unity Health Toronto, Canada
IT University of Copenhagen, Denmark
Olso Metropolitan University, Norway
University ofKentucky, USA

THE INORMS SCOPE FRAMEWORK–BRIDGING THE POLICY-PRACTICE DIVIDE

ABSTRACT

In 2001, the International Network of Research Management Societies' (INORMS) was formed to bring together research management societies and associations from across the globe. In recognition of the fact that research assessment was having a growing influence on the research management profession, INORMS established a Research Evaluation Group (INORMS REG) in 2018 to consider how best to ensure that research evaluation is meaningful, responsible, and effective. As part of the INORMS REG's aim of guiding university leaders and practitioners in the adoption and practice of responsible research evaluation, they developed a framework that both enabled evaluators to deliver on existing principles of responsible assessment and to address some additional critical elements (INORMS REG, 2021). As such, despite often being grouped together with other declarations or principles governing responsible research assessment like DORA, the Leiden Manifesto, and CoARA's ARRA, the SCOPE framework is actually a practical, step-by-step process by which such principles can be implemented.

The five stages of SCOPE are to:

- Start with what you value about the entity under evaluation;
- Consider the context of the evaluation;
- Explore all the options for evaluating;
- Probe deeply for any unintended consequences and discriminatory effects;
- Evaluate and evaluate your evaluation.

These stages are supported by three overarching principles, which are to:

- evaluate only where necessary;
- evaluate with the evaluated;
- evaluate using evaluation expertise.

Table 1 outlines how SCOPE seeks to deliver on some of the key elements of existing initiatives:

Principle/Declaration	Objectives	Relationship with SCOPE
San Francisco Declaration on Research Assessment (DORA), 2012 Seeks improved assessment of researchers and a better scholarly communication ecosystem.	18 recommendations for different stakeholders. The key themes are to: · Eliminate journal-based metrics. · Assess research on its own merits. · Take advantage of online publication possibilities.	SCOPE shares DORA's vision for better researcher assessment & eliminating the poor use of journal metrics, but is broader in focus, overseeing the responsible assessment of any entity.
Leiden Manifesto, 2015 Seeks more responsible use of bibliometrics in research assessment.	10 principles for the responsible use of bibliometrics in research assessment focused on: · Metrics supporting rather than supplanting expert assessment. · Mission-based performance assessment. · Accounting for variation by field in citation metrics.	SCOPE shares the Leiden Manifesto's vision for contextualized use of bibliometrics, but is not limited to quantitative indicators, as it accounts for qualitative measures too.
The Metric Tide, 2015 Seeks to guide a broad range of research assessment approaches.	Five principles for all forms of research assessment: · Robustness · Humility · Transparency · Diversity · Reflexivity	SCOPE also has a broad focus, but does not stop at providing principles, as it also provides a pragmatic, step-by-step process for evaluating responsibly that includes characteristics like value-led beginnings and a sense-checking probe stage.
Hong Kong Principles, 2020 Seeks to reward practices that lead to researcher integrity	Series of principles for assessing researchers that reward research integrity	SCOPE also offers value-based assessments, but does not prescribe what those values should be, instead letting the evaluators (together with the

Principle/Declaration	Objectives	Relationship with SCOPE
rather than unhelpful & limited publication-based rewards.	focused on: 1. Assessing responsible research practices 2. Valuing complete reporting 3. Rewarding the practice of open science 4. Acknowledging a broad range of research activities	evaluated) generate the values that are most meaningful to them.

Table 1. The relationship between SCOPE and other declarations and principles of responsible research assessment (RRA).

SCOPE bridges the gap between principles and their implementation by providing a structured and orderly framework by which evaluations can be designed and implemented as well as evaluated. Existing principles focus mainly on either evaluating a specific entity, like researchers in the case of DORA and Hong Kong Principles, or via a particular mechanism, like research metrics in the case of the Levin Manifesto and Metric Tide. SCOPE seeks to be applicable across the whole research ecosystem, enabling a responsible approach to evaluating any entity via any relevant mechanism.

However, SCOPE does not simply translate existing principles into practice but provides additional considerations not always addressed by principles of RRA, such as evaluating with the evaluated, evaluating only where necessary, starting with what is valued, and probing for unintended consequences. Thus, SCOPE, with its focus on implementation, plays a specific and novel role in the delivery of RRA.

Perhaps due to its simple, widely applicable, and pragmatic approach, the SCOPE framework is in wide usage by a range of organisations globally. Recent examples include the use of SCOPE by Indian funding agencies to develop new assessment mechanisms (Suchiradipta et al., 2023), by Colombian research professionals to develop a responsible assessment policy (Pallareset al., 2023), by Finnish research managers to create an institutional policy on assessment of researchers (University of Turku, n.d.), and by UK joint HE funding bodies in the redesign of the UK Research Excellence Framework (REF) (Himanen et al., 2023).

The features of SCOPE that have made it so attractive to a wide range of users also speak to some of the limitations of SCOPE as a framework. When given five seemingly simple steps, it is possible to just follow them mechanically, “tick the box," without giving enough consideration to what the steps are there for. Conversely, it may be difficult to discern when you have completed a step of the framework: How do you know when you have accurately identified your values? When have you prodded enough into unintended consequences? As a framework, SCOPE leaves this up to the discretion of the evaluating group, as time frames and amount of engagement may vary based on the organisation and topic being evaluated.

Another limitation of SCOPE is its lack of more concrete instructions on which tools or indicators should be used to ensure a responsible assessment. Such recommendations were beyond the scope of the framework, mainly due to the plethora of different approaches available and the complicated relationship they have with different contexts and purposes of evaluation. However, this can leave evaluators at a loss to know whether they have fully explored all of the options available to them.

Despite the inevitable limitations of any research evaluation framework that seeks to support assessors to design and critique evaluations in any setting, the traction that it has gained is evidence of the SCOPE Framework's usefulness and usability. The next step for SCOPE is to collect evidence, through further case studies, to support its own evaluation and development.

REFERENCES:

- Himanen L, et al.The SCOPE framework–implementing the ideals of responsible researchassessment [version 1; peer review: 2 approved withreservations].F1000Research2023,12:1241(https://doi.org/10.12688/f1000research.140810.1)
- INORMSREG(International Network Of Research Management Societies-Research EvaluationGroup):The SCOPE Framework: A five-stage process for evaluating researchresponsibly.International Network of Research Management Societies; 2021. (https://melbourne.figshare.com/articles/report/The_SCOPE_Framework/21919527/1)
- Pallares C, Chalela S, Tejada MA,et al.: The Colombian responsible metrics Project: Towards aColombian institutional, methodological instrument for research assessment.DORA Blog.2023
- Suchiradipta B, Koley M, Bharadwa J: Workshop on Research Assessment Practices in Indian FundingAgencies.Journal of Science Policy & Governance.2023;22(1).
- University of Turku:Policy for Responsible Assessment of Research and Researcher.University ofTurku; n.d.; 8.(https://zenodo.org/records/10037061)

Jürgen Janger, Kathrin Hofmann, Martin Unger Michael Ploder, Angelika Sauer
WIFO Austrian Institute of Economic Research, Austria
WIFO Austrian Institute of Economic Research, Austria
IHS, Institute for Advanced Studies, Austria
JOANNEUM RESEARCH Forschungsgesellschaft mbH, Austria
JOANNEUM RESEARCH Forschungsgesellschaft mbH, Austria

EVALUATING THE ECONOMIC AND SOCIETAL IMPACT OFGRANT-FUNDED BASICRESEARCH IN AUSTRIA

ABSTRACT

This contribution is based onastudy commissioned by the Austrian Science Fund FWF. Itmainlyfunds a variety of basic research projects, including single project funding for principal investigators,national and international research collaborations and career support schemes which also workthrough research projects, such as the START or theElise Richter programmes (Janger et al.,2019).Funding of PhD programmes is substantial(approx.16% of funding in 2017). Natural sciences,engineering and medicine,in whichapplications may arise sooner or more directly than in socialsciences and humanities, amounted to close to 80% of funding. While there have been evaluations oftheimpactofFWF funding schemes on research, little is known about their economic impact, at atime of growing pressure for research funders to show value for money(Dudley,2013)

Basic research is defined in the Frascati Manual as undertaken “without any particular application oruse in view”(OECD, 2015, p. 45).4Yet, many studies have empirically shown a positive impact ofacademic science or basic research on output indicators of innovative activity such as patents or oneconomic outcomes such as total factor productivity growth(Adams, 1990; Ahmadpoor & Jones,2017; Fleming et al., 2019; Jaffe, 1989; Mansfield, 1980, 1991, 1995). Others survey the various waysin which basic researchand its ensuing expanded stock of knowledgebecome economically useful toexplain this puzzle(Lane, 2009; Pavitt, 1991; Salter &Martin, 2001), leading scientists to be convincedof the crucial economic importance of basic research (e.g.,Dudley, 2013,p. 33:“History clearlyshows how fundamental science drives revolutions in technology”).

But tracing the economic impact of basic research(or fundamental knowledge)is notstraightforward.Many studies are doneforthe US, a large country atthe technological frontier witha large academic sector and a large share of high-tech industries, whichalleviatessome of themethodologicalissuesand increases the real effect of basic research: while the fundamental problemof a long time-lag betweenthe research and its application remains,internationalspilloversare lower.Economic impactwill notjust depend on the quality of the research, but also on the frameworkconditions for commercialising knowledge, such as the availability of venture capital(Lane, 2009).Research-intensive

sectors use basic research more intensively(Adams, 1990; Ahmadpoor & Jones,2017; Czarnitzki & Thorwarth, 2012). These difficulties make narrative impact approaches as in theUK Research Excellence Framework important(Khazragui & Hudson, 2015).

Simply taking US-based results as an indication for the economic impact of FWF-funded researchwould hence be problematic. In the following, we propose an approach that will show a variety ofsignificant economic impacts ofFWF-funded research, e.g. how it is used for inventive activity andhow it contributes to the economyand society.

We differentiate between two central ways to show the economicand societal impact of FWF-funded research:in the first way, the transmission channel between FWF funding and (potential)application is observable. Both statistical indicators of knowledge use and case-based narrativeapproaches can be employed ("stories").We collect data both from existing databases–such as Patstat, or Lens.org–and from a survey among FWF project leaders.In the second way, there is anunobservable relationship ("black box") between FWF-funded research results and economicoutcomes, but econometric estimation and modelling lead to "numbers", monetarised economicimpact. The first has the advantage of being easily understandable, allowing for the communication ofconcrete impacts. It does not allow though for the calculation of overall economic benefits in terms of value added or employment, a rate of return or elasticity with respect to FWF funding, which ispossible in the second way. The two approaches complement each other, the first one lending credibility to the second approach through concrete examples of the actual use of FWF-fundedresults and results from the first approach feeding into the second one.

Within the first approach("stories"–observed use of FWF-funded results/researchersin economic orsocietal applications), we analyse the quantity & quality of the following impacts:-

- Direct use of FWF-funded research results in patents by FWF-funded researchers, including relevance for grand challenges; via analysis of patents using Patstat database
- Use of FWF results in patents by non-FWF-funded researchers (overall contribution of FWF results to growth of technological knowledge, via patent-to-paper analysis, e.g., see Ahmadpoor & Jones, 2017,Munari et al., 2022), using the open source Lens.org database
- Contribution of FWF-funding to a skilled research workforce;via surveyresults
- Contribution-analysis based case studies on high-quality examples from the precedinganalysis,including start-ups, drugs, research tools&methods...;building on the literature, e.g. Mayne,2012

Within the second approach ("numbers"–overall economic impact), we analyse:

-Short-term impact through spending flows (PhD wages, material purchases) and any economicimpact of firms, drugs, etc. analysed in the first approach, using an input-output

model(Kratenaet al., 2017), building on recent literature pointing to the high business cycle stabilisation effectsof R&D support(Brautzsch et al., 2015)
-Medium-term impact of FWF-funded PhD qualifications on productivity, using the same model
-Longer-term impact of FWFfunding on productivity growth, GDP and employment, using suitable econometric methodologies

We capture the components of a production function/innovation chain, where FWF-funded researchis the input, outputs such as research publications and outcomes such as patents arecountable/observable, while wider economic impacts on overall productivity must beestimated/modelled. While we understand benefits of FWF funding in abroad way-codifiedresearch results are just one transmission channel next to human resources, spending flowsetc.ourresultsare likely tounderestimatethe true impact of FWFfunding,because impacts will arisein thefutureand somewhere else.We take the scientific quality of FWF-funded research projects as given–we do not evaluate FWF's selection processes.

REFERENCES:

- Adams, J. D. (1990). Fundamental Stocks of Knowledge and Productivity Growth.Journal of PoliticalEconomy,98(4), 673–702
- Ahmadpoor, M., & Jones, B. F. (2017). The dual frontier: Patented inventions and prior scientificadvance.Science,357(6351), 583–587.https://doi.org/10.1126/science.aam9527
- Brautzsch, H.-U., Günther, J., Loose, B., Ludwig, U., & Nulsch, N. (2015). Can R&D subsidiescounteract the economic crisis?–Macroeconomic effectsin Germany.Research Policy,44(3), 623–633.https://doi.org/10.1016/j.respol.2014.11.012
- Czarnitzki, D., & Thorwarth, S. (2012). Productivity effects of basic research in low-tech and high-tech industries.Research Policy.https://doi.org/10.1016/j.respol.2012.04.009
- Dudley, J. M. (2013). Defending basic research.Nature Photonics,7(5), 338–339.https://doi.org/10.1038/nphoton.2013.105
- Fleming, L.,Greene, H., Li, G., Marx, M., & Yao, D. (2019). Government-funded research increasinglyfuels innovation.Science,364(6446), 1139–1141.https://doi.org/10.1126/science.aaw2373
- Janger, J., Schmidt, N., & Strauss, A. (2019).International differences in basic research grant funding–asystematic comparison(Nr. 9-2019; Studien Zum Deutschen Innovationssystem). ExpertenkommissionForschung und Innovation (EFI).
- Khazragui, H., & Hudson, J. (2015). Measuring the benefits of university research: Impact and the REFin the UK.Research Evaluation,24(1), 51–62.https://doi.org/10.1093/reseval/rvu028

- Kratena, K., Sommer, M., Streicher, G., Salotti, S., & Valderas Jaramillo, J. M. (2017).FIDELIO 2:Overview and Theoretical Foundations of the Second Version of the Fully Interregional Dynamic EconometricLong-term Input-Output Model for the EU27. WIFO-EC Joint Research Centre, Institute forProspective Technological Studies.http://www.wifo.ac.at/wwa/pubid/61880
- Lane, J. (2009). Assessing the impact of science funding.Science,324(5932), 1273.Mansfield, E. (1980). Basic Research and Productivity Increase in Manufacturing.The AmericanEconomic Review,70(5), 863–873.
- Mansfield, E. (1991). Academic research and industrial innovation.Research Policy,20(1), 1–12.https://doi.org/10.1016/0048-7333(91)90080-A
- Mansfield, E. (1995). Academic Research Underlying Industrial Innovations: Sources, Characteristics,and Financing.The Review of Economics and Statistics,77(1), 55–65.
- Mayne, J. (2012). Contribution analysis: Coming of age?Evaluation,18(3), 270–280.
- Munari, F., Righi, H. M., Sobrero, M., Toschi, L., Leonardelli, E., Mainini, S., & Tonelli, S. (2022).Assessing the Influence of ERC-funded Research on Patented Inventions.OECD. (2015).Frascati Manual 2015. Guidelines for Collecting andReporting Data on Research andExperimental Development.
- OECD Publishing. Pavitt, K. (1991). What makes basic research economically useful? Research Policy,20(2), 109–119.https://doi.org/10.1016/0048-7333(91)90074-Z

Leonie van Drooge, Niki Vermeulen
LvD Impact & Evaluation, Netherlands
The University of Edinburgh / School of Social and Political Science, United Kingdom

EVALUATING THE ECONOMIC AND SOCIETAL IMPACT OFGRANT-FUNDED BASICRESEARCH IN AUSTRIA

ABSTRACT

How to evaluate strategic partnerships? International officers of six universities asked this simple yet challenging question. They collaborated in the Erasmus+ funded EVALUATE project . Researchers collaborate across borders and continents. Students go on exchange and go study abroad. Nothing new so far. Yet formalised international strategic partnerships between universities are more recent. These more formal agreements between universities include both research and education and cover a range of departments. As these partnerships are expected to contribute to strategic goals and have great impact, the question of evaluation becomes prominent.

However, it became clear that there is no precedent for the evaluation of strategic partnerships. Moreover, the term of strategic partnership has different meanings. Plus these partnerships are treated differently in each of the participating universities.

There is ample literature on internationalization, mobility, and environmental impacts of international collaboration. Yet it is not straightforward to find literature immediately relevant for international strategic partnerships between universities, nor on the evaluation of such agreements. Therefore, the notion of international strategic partnerships was first unpacked, to understand the different forms and formats it can take. This approach delivered keywords for a broader literature review, which put forward various, separate bodies of literature that are all relevant to international partnerships. The literature review provided our evidence base to develop the framework. (EVALUATE project 2022, 112-152).

The project team decided to develop a framework from scratch. It chose a co-creative approach to make sure that the framework was embedded in both theory and practice. Academics of two of the participating universities guided staff in international offices of the six universities through the evaluation of a specific strategic partnership. They then iteratively developed the framework, and adjusted it based on the feedback and responses of the international office staff. Only through mutual exploration of the practice of international strategic partnerships, did it become possible to find ways to mobilise existing insights from the evaluation literature and integrate this in a meaningful way. As such, the creation of an evaluation framework for strategic partnerships between universities was an experiment in co-creation, developing evaluation in practice with those using the framework. By working on the development of the framework from the start, it

was ensured that the resulting framework fits existing practices and that it can more easily be integrated into ongoing partnership work. This has enhanced the capability for implementation and allowed project participants to communicate results and spread evaluative thinking.

The EVALUATE project has been a learning journey for all. The governance philosophy of the various universities, the partnership activities, the goals of the partnerships and the relation to university strategies differ between the partners and cases. The lack of consistent and aligned evaluation practices was confirmed throughout the project. Consequently, partners realised that a rigid framework with clear measures or benchmarks is not realistic or useful. Evaluation is best integrated from the start to the end of partnerships. If integrated well, evaluation is a cyclical activity returning in every phase of the partnership, underpinning decisions, and new actions.

The project to co-create an evaluation framework was innovative in a number of ways. Firstly, the framework provides a solution for university staff working in the areas of partnership development and evaluation. This is an area of need, given recent and current prioritisation of international partnerships in both institutional, national and international strategies that concern university education and research. In addition, the co-creation with a project team consisting of academic evaluation specialists and international officers was key. The combination of academic and professional inquiry enabled the project to apply robust academic methodologies and scrutiny alongside professional experience and expertise in the field. In addition, the project was complimentary to a wide range of other initiatives and can support the higher education sector to forge greater understanding of the impact of university partnerships. For example, the UK Research Excellence Framework (REF) highly values the impact of research. The impact of research often happens through partnerships – and this project provides a framework to understand the value of partnerships.

Perhaps the most interesting chapter of the handbook is Lessons Learnt from the EVALUATE project, written by the international officers (EVALUATE project 2022, p. 41-46). While these officers initially struggled with the diverse perspectives, they now acknowledge the variety and suggest taking these into account and use participative methods. They provide recommendations on such diverse topics as data collection, the use of evaluation and the implementation of strategic partnerships. They question unrealistic expectations and bold claims, such as turning students into “global citizens” or using research to address “global challenges”. And they advise thinking about evaluation as an opportunity to build capacity and involve and inspire partners. They also characterise their own process of change in respect to evaluation, from first awakening to building awareness to maturing appreciation. The co-creation approach fits with current calls for reforming evaluation culture (CoARA, 2022). First of all, the project steered away from purely quantitative approaches. Although some universities had a good overview of investments on the one hand, and outputs in terms of publications and funding on the other hand, it was agreed that good evaluation requires more than measuring what can be quantified. At the start, the evaluative cycle was introduced, that shows how good evaluation practice is integrated throughout the life cycle of a partnership, and that a range of methods can be used to answer evaluative questions, developing mixed-method approaches (Better Evaluation 2023). In line with this, the project connected to current shifts in evaluation culture, moving away from accountability towards a formative role for evaluation (Molas-Gallart et al. 2021, Dinges et al. 2020). Moreover, it was stimulated to evaluate with the partner. It meant a shift from evaluation for accountability towards evaluation as communication device and mutual learning between partners. (Spaapen 2015, Joly and Matt 2022). As such, the development of the framework contributes to novel approaches to evaluation policy that challenge current practices. In fact, the process approach can be viewed as an experiment, and it can inspire the development of other types of frameworks and contribute to policy learning.

Ulrike Felt, Maximilian Fochler
University of Vienna, Austria

HIRING EXCELLENCE: BETWEEN IDEAL VALUE REGISTERS AND QUALITY ASSESSMENT IN PRACTICE

ABSTRACT

The pursuit of quality in academic performance within higher education institutions (HEIs) is fundamental to their mission and identity. This pursuit is particularly critical during appointment processes of full professors, pivotal moments that shape the medium-term trajectory of these institutions. However, defining and measuring quality poses significant challenges, especially amidst rapid transformation and increased societal engagement within HEIs. Expansion of HEIs' activities, such as third mission initiatives and open science, has diversified the dimensions of quality, complicating the articulation and weighting of these dimensions in appointment processes. Balancing factors like research, teaching, and societal impact is increasingly complex, requiring careful consideration in candidate selection. Moreover, quality assessment within individual dimensions like research and third mission activities can be contentious. International competition for talent further heightens the pressure to make decisions swiftly amid complexity, fostering a desire for easily quantifiable standards.

It's crucial to recognize that quality criteria in appointment processes shape researchers' long-term self-perception as well as institutional culture. These criteria influence which activities are valued within academia, impacting researchers' career trajectories and junior researchers' career aspirations. Understanding how quality definitions in appointments shape institutional culture and resource allocation is paramount.

We argue that analyzing quality negotiation in concrete processes like appointment procedures provides a realistic assessment, moving beyond abstract debates. Austrian universities, recognizing appointment processes' significance in strategic development, devote considerable attention to them. Our study aims to map the ongoing discourse on defining and measuring academic performance quality in Austrian HEIs' appointment procedures. Focusing on appointments to university professorships, the study seeks to understand perspectives on future challenges and opportunities. By examining the intricacies of quality assessment in appointment processes, the study aims to inform strategic decision-making within Austrian HEIs. This research acknowledges the complexity of quality determination in academia and seeks to provide actionable insights for enhancing institutional effectiveness and fostering academic excellence.

The study began by requesting internal documents from participating institutions regarding their appointment processes and quality assurance procedures in accordance

with the respective Universities Act. These documents underwent initial analysis to understand the procedures and develop interview questions.

Semi-structured qualitative interviews were then conducted with experts from various positions within the institutions, such as the Senate Office, Quality Assurance, and the Rectorate Office. These interviews aimed to comprehend the process logic of appointment procedures and assess quality dimensions within the procedures. Subsequently, interviews were conducted with members of the rectorate responsible for appointment procedures to understand strategic considerations and management visions, as well as challenges from a management perspective. Additionally, heads of faculties were interviewed to capture management and process perspectives and explore potential differences in quality perspectives based on disciplinary cultures.

All interviews were transcribed with confidentiality assured, and a thematic analysis was conducted on the collected material, guided by the study's framework questions. Themes were identified and categorized, forming the basis of the results presented in the report. Furthermore, we developed graphical representations of the hiring choreographies. This comprehensive approach, combining document analysis and interviews, facilitated a structured yet flexible examination of appointment processes and quality assessment within Austrian HEIs, allowing for a nuanced understanding of the subject matter. We aimed at capturing the processes and participating key actors involved in defining quality and ensuring quality assurance in appointment procedures, the translation of quality concepts into the appointment processes, the relevant dimensions of quality and their visibility in the procedure, and the interrelationship between different quality dimensions.

Even though all Austrian HEIs are governed by the same legal framework, we observed substantial differences when it comes to the concrete practices of quality assessment procedures in hiring procedures of full professors. Following we will identify some key angles which need consideration to reconcile more abstract value registers and hiring practices.

- Engaging in meaningful broader discourse about quality within the institution provides a valuable framework for appointment processes, aiding the appointment committee's decision-making. Establishing consistent quality standards across various processes, including appointments, individual evaluations, and evaluations of subunits, is essential. Additionally, collectively developing a narrative that aligns quality concepts with the organization's strategic mission can guide decision-making effectively.

- Quality assurance in appointment procedures relies on two kinds of expertise from involved stakeholders – procedural knowledge and evaluation expertise. Acknowledging and cultivating both is crucial for fostering a broader quality culture.
- Designing formal procedures thoughtfully shapes how quality is assessed. Administrative details, such as structuring application documents, influence the committee's ability to assess quality dimensions. Planning hearings needs more attention to the mix of formal and informal elements and their function in the process.
- Explicit discussion of quality criteria within the committee fosters strategic discourse. Defining criteria early in the process and in sufficient detail promotes transparency and prevents strategic adjustments.
- A prior strategic profiling can help the commission to decide, e.g., how to weight quality dimensions and how to deal with the relationship between past performance and future potential in the process.
- Quality assessments in appointment procedures must avoid over-reliance on any single dimension (e.g., research) or type of evidence (e.g., quantitative indicators) to ensure a balanced outcome.
- Teaching and didactics are increasingly recognized as important quality dimensions. Active student involvement in evidence design, such as teaching samples, and structuring application documents for comparative evaluation of didactic competence are pivotal.
- The third mission's growing importance as a quality aspect raises questions about its separate management or integration within research and teaching dimensions.
- Leadership competence assessment lacks alignment between perceived importance and evaluation quality. Providing better evidence, such as through structured application documents and involvement of HR consulting experts, could improve assessment.
- Addressing the social fit of candidates is sensitive yet vital. Explicitly defining its role as a secondary criterion and assessing it in later phases of the procedure, possibly through informal elements in hearings, is advisable.
- Despite its international significance, open science is not widely recognized as a quality dimension in surveyed universities. Increasing attention to this topic in institutional quality discourses is recommended.

Gaston Heimeriks, Anna J. Wieczorek, Orlando Martin, Vazquez Villegas, Oscar Yandy Romero-Goyeneche, Jonas Torrens, Anna Shindler
Utrecht University, Netherlands
Eindhoven University of Technology, Netherlands
Eindhoven University of Technology, Netherlands
Eindhoven University of Technology, Netherlands
Utrecht University, Netherlands
Utrecht University, Netherlands
Unknown Istitution

TRANSFORMATIVE RESEARCH AT TECHNICAL UNIVERSITIES

ABSTRACT
Universities are presently grappling with how to address complex and rapidly worsening socioenvironmental challenges of recent decades. In the Netherlands, a growing number of proposals for reframing universities' research efforts around 'Transformative research' (TR) have emerged. This 'transformative' turn builds on similar discussions concerning innovation policy and could indicate an emerging realignment between research activities and the societal response to rapidly unfolding and compounding. Nevertheless, how Transformative Research is actually taking root remains fuzzy and poorly supported by empirical evidence, particularly in technical universities that are strongly tied to more traditional conceptions of engineering and natural sciences research. This paper seeks to uncover whether and how transformative research is emerging in the research cultures, structures, and practices at the Eindhoven University of Technology (TU/e).

In particular, we sought to understand how researchers and communities committed to transformative research frame their activities and whether the university's institutional environment supports this emergent approach. In this paper, we researched:

- (How) does the recent transformative research debate differ from previous conceptualisations, and how does it implicate technical universities?
- How is transformative research for sustainability emerging or apparent in the research cultures, structures, and practices at the TU/e?
- How can the emerging transformative research and existing engineering and sciences research interact in more productive ways?

Increasingly, European universities are grappling to embed sustainability in their strategies to proactively respond to societal challenges and the climate crisis. For some, this is a matter of aligning with new rationales for science, technology, and innovation policy (Schot and Steinmuller, 2018, Parker and Lundgren, 2022) and find ways to contribute to societal missions. For others, it is a reckoning with a long period of academic productivism, tied to an excessive emphasis on publications in detriment of other forms of societal

engagement. Across the sector, universities are launching new initiatives to address the mounting societal and ecological challenges, and beginning to recognise that their current model is need of renewal and transformation (Fazey et al., 2021; Loorbach and Wittmayer, 2023; Reed and Fazey, 2021).

In this context, the term 'transformative research' is emerging as an overarching rationale for research practices that respond to these challenges (e.g., Schneidewind et al., 2016). Among Dutch universities, this notion has started to be explored and operationalised, primarily by researchers already active in sustainability transitions and transformations research (Chambers et al. 2020, Wittmayer et al. 2018, 2021). Those debates show promising signs, but much remains to be clarified about how transformative research is already taking root at universities, outside dedicated programmes. Significantly, technical universities that are usually very engaged in pursuing technological solutions for existing challenges have not received dedicated attention.

The credibility cycle of science (Latour and Woolgar, 1986) provides a theoretical lens for understanding the opportunities and constraints of transformative research. It draws attention to the different mechanisms that legitimize and incentivize scientists and scientific projects in different, interrelated areas of knowledge production including providing data, data analyses, equipment, research design, dissemination of ideas and obtaining resources. Each step of the cycle is influenced by distinct institutional challenges. These institutional challenges inform and influence research projects and the credibility cycle helps us understand how scientific motivations and behaviours are institutionalised in various ways (e.g. funding programmes, peer review, career progression, conferences).

To better understand the present status of TR at TU/e, we employed a mixed method comprising a bibliometric and a qualitative component. The former aimed to help us select research communities and individuals likely to pursue some form of (transformative) sustainability research. With the latter, we explore how these communities frame and justify their efforts in the context of the credibility cycle. First, we used bibliometric analysis to map knowledge clusters as a proxy of knowledge communities. Using the university's research publications and a classification procedure based on the Sustainable Development Goals (SDGs), we identify researchers that can potentially contribute to developing transformative research. The methodology is based on Goyeneche et al. (2022) and Schot et al. (2018), who proposed that knowledge communities integrating knowledge from multiple SDGs can better contribute to implementing transformative science. We selected four communities with contrasting characteristics (most dissimilar) and interviewed a selection of their researchers to understand how they frame their contributions. We coded these interviews to compare the communities' goals, rationales, interactions, and implementation approaches.

Our results show that TR communities orient themselves in ways broadly compatible with the framing of transformative research emerging in the literature, albeit with diverse focal points, emphasis, and approach to alternatives and the emphasis on technological solutions. Across all communities studied, the rationale mentioned combinations of urgency, complexity, justice, diversity, and curiosity. Transformative research did not stand out as an entirely new goal but rather as a complementary orientation that draws from and extends other modalities and goals.

Our approach offers a more systematic understanding of the development of transformative research areas. Identifying and understanding the underlying conditions that enable and constrain the emergence of transformative research can help to increase the transformative potential of scientific knowledge. Funding for transformative research, to name one example, is not always straightforward as it often interdisciplinary and complex.

REFERENCES:

- Fazey, I., Hughes, C., Schäpke, N.A., Leicester, G., Eyre, L., Goldstein, B.E., Hodgson, A., MasonJones, A.J., Moser, S.C., Sharpe, B., Reed, M.S., 2021. Renewing Universities in Our Climate Emergency: Stewarding System Change and Transformation. Front. Sustain. 2, 677904. https://doi.org/10.3389/frsus.2021.677904
- Fazey, I., Schäpke, N., Caniglia, G., Patterson, J., Hultman, J., Van Mierlo, B., Säwe, F., Wiek, A., Wittmayer, J., Aldunce, P., Al Waer, H., Battacharya, N., Bradbury, H., Carmen, E., Colvin, J., Cvitanovic, C., D'Souza, M., Gopel, M., Goldstein, B., Hämäläinen, T., Harper, G., Henfry, T., Hodgson, A., Howden, M.S., Kerr, A., Klaes, M., Lyon, C., Midgley, G., Moser, S., Mukherjee, N., Müller, K., O'Brien, K., O'Connell, D.A., Olsson, P., Page, G., Reed, M.S., Searle, B., Silvestri, G., Spaiser, V., Strasser, T., Tschakert, P., Uribe-Calvo, N., Waddell, S., Rao-Williams, J., Wise, R., Wolstenholme, R., Woods, M., Wyborn, C., 2018. Ten essentials for action-oriented and second order energy transitions, transformations and climate change research. Energy Res. Soc. Sci. 40, 54– 70. https://doi.org/10.1016/j.erss.2017.11.026
- Gravem, S.A., Bachhuber, S.M., Fulton-Bennett, H.K., Randell, Z.H., Rickborn, A.J., Sullivan, J.M., Menge, B.A., 2017. Transformative Research Is Not Easily Predicted. Trends Ecol. Evol. 32, 825–834. https://doi.org/10.1016/j.tree.2017.08.012
- Institut, W., n.d. Transformative Research - Wuppertal Institute for Climate, Environment and Energy [WWW Document]. URL https://wupperinst.org/en/research/transformative-research (accessed 10.18.23)

- Loorbach, D.A., Wittmayer, J., 2023. Transforming universities: Mobilizing research and education for sustainability transitions at Erasmus University Rotterdam, The Netherlands. Sustain. Sci. https://doi.org/10.1007/s11625-023-01335-y
- Meisch, S., 2020. Transformative Research. The IASS Approach 28 pages, 564 KB. https://doi.org/10.2312/IASS.2020.021
- NSFB, 2007. Enhancing support of transformative research at the National Science Foundation. National Science Foundation Board.
- Reed, M.S., Fazey, I., 2021. Impact Culture: Transforming How Universities Tackle Twenty First Century Challenges. Front. Sustain. 2, 662296. https://doi.org/10.3389/frsus.2021.662296
- Tilbury, P.D., 2011. Higher Education for Sustainability: A Global Overview of Commitment and Progress. High. Educ.
- Trevors, J.T., Pollack, G.H., Saier, M.H., Masson, L., 2012. Transformative research: definitions, approaches and consequences. Theory Biosci. 131, 117–123. https://doi.org/10.1007/s12064-012- 0154-3
- WGBU (Ed.), 2011. World in transition: a social contract for sustainability. German Advisory Council on Global Change, Berlin.
- Wittmayer, J.M., Hölscher, K., Wunder, S., Veehoff, S., n.d. Transformation research: exploring methods for an emerging research field (No. 01/2018). Environmental Research of the Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety, Dessau-Roßlau.

Janne Pölönen, Reetta Muhonen, Mira Söderman, Maria Pietila, Tiina Sipola, Laura Niemi, Anna-Kaisa Hyrkkanen
Federation of Finnish Learned Societies , Finland
Tampere University, Finland
Federation of Finnish Learned Societies, Finland
University of Eastern Finland, Finland
Federation of Finnish Learned Societies, Finland
University of Turku, Finland
Aalto University, Finland

RESEARCHERS' VIEWS ON DIVERSITY OF CAREER ASSESSMENT CRITERIA IN FINLAND

ABSTRACT

One of the main goals of the responsible research assessment agenda is to move away from a narrow set of assessment criteria to recognize the range of contributions researchers make to science and society. The first commitment of the Agreement on Reforming Research Assessment (ARRA), signed already by hundreds of organisations, is to "recognise the diversity of contributions to, and careers in, research in accordance with the needs and nature of the research" (CoARA, 2022). ARRA builds on an increasing number of international and national initiatives, which over the past decade have called for diversification of criteria for recruitment, promotion and funding of researchers: for example DORA declaration, the Leiden manifesto, The Metric Tide, Room for Everyone's Talent, Helsinki Initiative, SCOPE Framework, and the Hong Kong Principles (Curry et al, 2020; Peruginelli & Pˆlˆnen, 2023).

The European Commission has promoted the recognition of diversity in career assessment since 2005 through the European Charter for Researchers (European Commission, 2005; Council of the European Union, 2023). To promote more holistic approach, the European Commission's Open Science Policy Platform (OSPP) formulated in 2017 the Open Science Career Assessment Matrix (OS-CAM), which provides 42 possible Open Science assessment criteria across six main areas of activities: research output; research process; service and leadership; research impact; teaching and supervision; professional experience (European Commission, 2017). The European Competence Framework for Researchers (ResearchComp) outlines 38 competencies across 7 competence areas (European Commission, 2022).

OS-CAM approach has inspired the development of more comprehensive career assessment tools and matrices at institutional and national level. NOR-CAM: A toolbox for recognition and rewards in academic careers in Norway (Universities Norway, 2021) provides examples of results and competencies across six areas of academic work, with

indication of their preferred documentation and self-reflection. Already two Finnish universities (Turku and Oulu) have developed their own CAMs based on the model of OS-CAM and NOR-CAM. Also new versions of CAM have been developed for diverse career profiles and/or stages (Pain, 2023).

In Finland, a broad-based working group set up by the Federation of Finnish Learned Societies (TSV) in 2018 produced a national recommendation Good Practice in Researcher Evaluation in Finland (Working group, 2020). A working group appointed by the Steering Group for Responsible Evaluation of the Researcher is currently planning the Finnish Career Assessment Matrix (FIN-CAM) to support a comprehensive, systematic and transparent documentation and consideration of the different areas of academic work. Building a preliminary draft of FIN-CAM, the working group wanted to engage the Finnish research community by means of a survey with the aim to better understand how the researchers would like to be evaluated across areas of assessment, research output, research process, teaching, societal impact, leadership and open science.

The survey consisting of 31 closed and open-ended questions was carried out in November-December 2023 by TSV. It was disseminated by TSV through social media and newsletter to almost 300 member societies, by the Finnish Union of University Researchers and Teachers (FUURT) and the Finnish Union of University Professors to their members, by the rectors conferences for universities (UNIFI) and universities of applied sciences (ARENE) and by network of State Research Institutes (TULANET) to their member organisations.

The added value of the survey is to provide a researcher perspective on the diversity of careerassessment criteria. Researchers were defined broadly as persons who participate in various ways in professional activities aimed at producing new knowledge, applying knowledge in a new way, developing knowledge or making use of it in education. Also doctoral researchers were included in the target group. The survey received 440 responses from researchers representing a wide range of disciplines, research organisations and career stages.

During their careers, most respondents had faced evaluation situations where they felt inappropriate methods or criteria had been used (55%), or where the goals, criteria, or data of the evaluation had not been openly known to all parties (52%). In open answers, one key challenge was framed by a respondent as follows: “the process is not actually carried out with the aim of genuinely evaluating competence, but rather it is mainly a matter of theatre, where the aim is to find justifications for the outcome desired in advance.” Clarity, transparency and consistent application of evaluation criteria emerged as a good practice. Narrow focus on research and especially publication and venue based metrics - and their priorisation over other valuable contributions to science and society - was pointed out as a major shortcomings encountered in assessments.

űRespondents were presented a series of closed questions about the importance of criteria across different areas of academic work in the following format: “If you could decide how your qualifications or performance as a researcher were evaluated, which of the elements listed below would you consider to be important?” They were asked to rate the importance of listed elements from their personal perspective on the following scale: “very important”, “moderately important”, “should not be taken into account in evaluation”, “cannot say”. Most respondents regarded a great variety of experience, competences and contributions related to outputs, research process, teaching, societal interaction, leadership and open science as very important or moderately important for their assessment. On average, 440 respondents considered 26 out of 109 elements listed in the survey very important. In the presentation we will specifically compare the views of researchers from STEM and SSH fields.

Findings from the survey provide strong support for the responsible research assessment agenda (e.g. CoARA) and the development of the FIN-CAM. Together with a narrative CV and research information systems, FIN-CAM can support the production of structured and evidence-based information on diverse contributions for responsible researcher assessment.

REFERENCES:

- CoARA - Coalition for Advancing Research Assessment (2022). Agreement on Reforming Research Assessment. https://coara.eu/app/uploads/2022/09/2022_07_19_rra_agreement_final.pdf
- Curry S., de Rijcke, S., Hatch, A., Pillay, D. van der Weijden I., Wilsdon, J. (2020). The Changing Role of Funders in Responsible Research Assessment. Research on Research Institute (RoRI). https://eprints.whiterose.ac.uk/171602/1/GRC_Responsible_Research_Assessment_RoRI_WP03_finalpd f.pdf
- European Commission (2005). The European Charter for Researchers. https://euraxess.ec.europa.eu/sites/default/files/brochures/am509774cee_en_e4.pdf

- European Commission (2017). Evaluation of Research Careers Fully Acknowledging Open Science Practices. Rewards, Incentives and/or Recognition for Researchers Practicing Open Science. https://data.europa.eu/doi/10.2777/75255
- European Commission (2022). Knowledge ecosystems in the new ERA – Using a competence-based approach for career development in academia and beyond. https://data.europa.eu/doi/10.2777/150763
- Council of the European Union (2023). Council conclusions on a European framework to attract and retain research, innovation and entrepreneurial talents in Europe. https://eur-lex.europa.eu/legalcontent/EN/TXT/PDF/?uri=OJ:C_202301640
- Pain, E. (2023). How academia is exploring new approaches for evaluating researchers. Science. https://doi.org/10.1126/science.caredit.adj4869
- Peruginelli, G. & PˆIˆnen, J. (2023). The legal foundation of responsible research assessment: An overview on European Union and Italy, Research Evaluation, 32:4, 670–682. https://doi.org/10.1093/reseval/rvad035
- Universities Norway (2021). NOR-CAM: A Toolbox for Recognition and Rewards in Academic Careers. https://www.uhr.no/en/_f/p3/i86e9ec84-3b3d-48ce-8167-bbae0f507ce8/nor-cam-a-tool-box-forassessment-and-rewards.pdf
- Working Group for Responsible Evaluation of a Researcher (2020) Good Practice in Researcher Evaluation. Recommendation for the Responsible Evaluation of a Researcher in Finland. https://avointiede.fi/sites/default/files/2020-03/responsible-evalution.pdf

Silvia Dobre, Andrew Plume
Elsevier, United Kingdom

SIZING THE MOVEMENT: THE TRAJECTORY OF RESEARCH CULTURE IN THE UK

ABSTRACT

Research culture is a subject intensely discussed and debated in the UK. The commonly adopted definition was published by Royal Society in January 2018: "Research culture encompasses the behaviours, values, expectations, attitudes and norms of our research communities. It influences researchers' career paths and determines the way that research is conducted and communicated". In the past 6 years the conversations, events and initiatives around the topic of research culture have increased considerably in the UK. The purpose of this study is to identify the trajectory of the main initiatives in this area, the people contributing to it and the ways in which this has been done.

Most of the data is openly available and was collected from the websites of the different stakeholders operating the research culture area. In some cases, the data was enhanced with further information available on LinkedIn. The analysis of the research papers published in this area used the Scopus data.

The redesign of the UK's national research assessment exercise (REF) reshapes the incentives within the research system and highlights the importance of research culture in underpinning excellent research with wider social impacts. Aiming to "reward those institutions that strive to create a positive research culture", REF2029 allocates a 25% weighting to the People, Culture and Environment (PCE) component.

The poster presents the diversity of the stakeholders involved in the research culture discussions in the UK and it captures the initiatives which have been witnessed in the past years. Although articles, publications and events in this area started to emerge 7 years ago, their spread, depth and variety has increased considerably in the most recent years. We started by analysing the articles published in the field of research culture and the contribution of UK authors. The analysis identifies trends in the topics associated with this area and new directions for research in this field. The poster also captures the way in which the higher education sector is moving in reaction to the changes in REF and the increased importance of the research culture agenda. We are witnessing a flurry of jobs with "research culture" in the title that have been advertised by the UK universities in the past 18 months, although a substantial proportion of them are fixed-term roles, linked with the preparation of the REF submission. Although this is ironic, as addressing precarity is part of building a positive research culture, it reflects the UK universities' stretched resources and the fact that some of these positions are recruited as part of an institutional funding for research culture award.

The analysis of the data related to the number of people working in research culture jobs in the UK higher education sector paints a very interesting picture. Some institutions are very advanced, while others are making the first tentative steps. The people working in these newly created roles (based in Research Services / Research and Innovation / Research and Enterprise) were mostly recruited internally and there is very little influx from outside HE sector. A big proportion of those employed in higher-grade roles are educated to the PhD or Master level, but overall there is very little ethnic diversity. Variations in the teams' size is correlated with the type of institution and geographical location.

The first International Research Culture Conference hosted in 2023 at the University of Warwick identified 12 topics related to research culture. Alongside open research, improving reproducibility, research integrity there were also included bullying and harassment, researcher recruitment, career development, researcher training, research leadership. We analyse which are the themes that universities are focusing on, what initiatives, policies, activities and events are at the core of the research culture and how open are these discussions. The rise of the research culture in UK is parallel to and linked with the rise of the open research agenda. Over the recent years we have seen the shift from one-day annual Open Research and Research Culture events organised for an internal institutional audience to conferences open to the entire sector such as the Research Integrity and Culture online events or the Open Research Weeks organised by a group of universities. A National Centre for Research Culture was set up last year by the University of Warwick and this spring we will see a research culture special issue of the Exchanges - The Interdisciplinary Research Journal.

The forthcoming results of the Technopolis and CRAC-Vitae project, aiming to develop a shortlist of indicators to be used to evidence and support institutions' PCE submissions as part of a structured questionnaire for REF2029 submissions will, no doubt, bring more clarity and start other debates in UK and in the field of responsible research assessment. There are parallel and interlinked initiatives going on elsewhere in the world that connect with this UK specific conversation, such as the CoARA early- and mid-career researchers working group intensely focused on assessment and research culture. This poster will offer the opportunity for interesting conversations and debates on the UK perspective in the wider context of an international conference which brings together researchers from other countries and early career researchers for which this topic is particularly relevant.

Historically "research culture" wasn't an established phrase and universities and other institutions occasionally looked into the "organisational culture" as part of their Human Resources activity. Until a few years ago there was no one in post with a "research culture" job title, there weren't conferences, webinars, discussions or research papers in this field. This poster looks at where research culture originated, how it grew and evolved, alongside the timeline of the sector's initiatives. It emphasises how UK universities have quickly moved to embrace and develop the research culture agenda, sizing the movement and riding the tide of change. The implications of the progress in this field in UK are highly relevant for the international conference on science, technology and innovation indicators; we see the need and the strong appetite to develop indicators able to assess the effectiveness of interventions aimed at improving research culture.

Lottie Provost, Giulia Malaguernara, Zenia Xenou, Angeliki Tzouganatou
Consiglio Nazionale delle Ricerche, Italy
Open AIRE, Greece

REFRAMING RESEARCH ASSESSMENT: TOWARDS A COMPREHENSIVE FRAMEWORK FOR RESEARCHER PROFILES

ABSTRACT

In recent years, the European Research Area policy agenda has prominently placed the reform of the research assessment system at the top of its policy actions, recognising that the way research projects, researchers, research units, and research institutions are assessed is fundamental for a well-functioning Research and Innovation system. Policy efforts have shifted to focus on accelerating the move beyond traditional, publication-based, research assessment methods, highlighting their inappropriateness, especially with regard to the increasingly collaborative and interdisciplinary nature of research (European Research Area policy agenda, 2022). There is indeed consolidated evidence that publication-based metrics such as the Journal Impact Factor and the h-index fail to reflect the broad range of activities that make up research, and are widely (mis)used as proxies for assessing the quality, performance and impact of research and researchers (Institut de France, 2011; Hicks et al., 2015; Pontika et al., 2022; DORA, 2024). Critics of the system have also shed light on the ways in which the current assessment system has created perverse incentives for researchers to prioritise aspects such as publication venue and number of citations (Edwards et al., 2017), leaving aside essential aspects of scientific knowledge production such as research quality, collaborative open research methods, and the impact of research on society (Di Donato, 2024). In response to the identified challenges, the European Commission has driven the movement for reform, seeking to establish a clear and common direction for the advancement of research assessment practices. In 2021, the European Commission Scoping Report "Towards a reform of the research assessment system" (European Commission, 2021) called for research proposals, researchers, research units and research institutions to be "evaluated on their intrinsic merits and performance rather than on the number of publications and where they are published, promoting qualitative judgement with peer-review, supported by responsible use of quantitative indicators." Echoing this call, signatories of the Agreement on Reforming Research Assessment (ARRA) (CoARA, 2022), who are now over 700, have undertaken to uphold a series of commitments, including the recognition of the diversity of contributions careers in research, and to base research assessment primarily on qualitative evaluation for which peer review is central, supported by responsible use of quantitative indicators. However, for most ARRA signatory organisations, implementing such changes remains a challenge. In particular, tailoring research assessment practices to different disciplines, career stages

and research outputs further increases this challenge and the lack of a high-quality and open infrastructure appears to be a major obstacle.

Innovative tools services to support organisations in the implementation of Open Science-aware RRA

A number of EU-funded projects are currently tasked with supporting the ongoing policy reforms and designing new ways to incentivise higher quality research, collaboration and Open Science practices (European Commision, 2024).

Among these, the Horizon Europe project GraspOS: Next Generation Research Assessment to Promote OpenScience addresses the need for new services and tools to support a research assessment system that incentivises Open Science practices. The project aims to develop a data infrastructure facilitating qualitative and quantitative assessments, ultimately supporting the practical implementation of the reform at various levels and fostering the adoption of an Open Science-aware responsible research assessment. In order to promote a comprehensive and flexible evaluation framework which accommodates for the diversity of research disciplines and practitioner backgrounds, the project focuses on various levels of research assessment: national , institutional , and individual.

In this context, GraspOS is working on the development of an innovative tool aimed at guiding research funding and performing organisations in the implementation of the ARRA commitments, and at supporting researchers in providing a better, more comprehensive view of their contributions to science and society. The tool is envisaged as a framework for Researcher Profiles whose design will follow closely the latest policy recommendations and guidelines promoting a responsible approach to research assessment (INORMS Research Evaluation Group, 2023; DORA, 2024) with the aim to provide a wide-ranging and flexible framework which supports a research culture that values diverse practices, and a research environment that values comprehensive quality and societal impact of research over mere numerical output. By combining quantitative information on research activities and outputs with an alternative qualitative perspective, the researcher profile we are developing should help provide a broader understanding of a researcher's values and the broader influence of their work.

Designing the Framework for the Researcher Profile: Methodology overview

The framework for the Researcher Profile aims to provide a customisable tool that allows researchers to showcase their diverse contributions to research, knowledge and innovation and which will be interoperable with ORCID and the OpenAIRE Graph, one of the largest databases collecting and linking different types of research contributions.

The design of the framework started with a landscape analysis of current services providing indicators on research activities, confirming that most indicators focus on scientific publications 4 . To consider which novel indicators to include in the framework for the Researcher Profile, our efforts will build on the ongoing work of Horizon Europe projects OPUS and PathOS. The OPUS project is working on a framework to assess researchers including Open Science dimensions to ensure that such practices are explicitly recognised and rewarded (O'Neill, 2023). The PathOS project has published a first version of the Open Science Indicator Handbook, which aims to provide guidance on how to make use of a wide range of Open Science indicators. The efforts undertaken by both projects will be of fundamental importance in designing a comprehensive framework for the Researcher Profile in which the scope of Open Science is not limited to indicators relating to Open Access publishing, but is rather considered in its most encompassing definition, which can be found in the UNESCO Recommendation on Open Science (UNESCO, 2021).

An innovative aspect of the framework will be the inclusion of an alternative Narrative CV section, displayed as an interactive timeline within the Researcher Profile. In this timeline, researchers will be able to inform evaluators about their main activities, highlighting the research areas they contributed to and the influence of their work in the field of science, but also in education, policy, economy and for society at large. This approach supports evidence-based Narrative CVs by combining qualitative and quantitative information in an interoperable and non-commercial service which will be provided by OpenAIRE. This novel approach aims to establish a framework of reference in which researchers can select the activities and contributions they wish to highlight.

Finally, the framework will be refined in collaboration with the nine GraspOS Pilots who each represent a specific context in the research assessment system (National research funding and performing organisations, CRIS systems, universities and university departments, disciplines). The pilots will provide practical feedback on the suggested components of the framework.

Recommendations on the responsible use of Open Science indicators in the development of tools and services

As with any new measure, the design and development of this framework should carefully take into account a variety of potential pitfalls. The GraspOS project aims to balance the quantitative indicators with qualitative assessment. However, there may be a risk that specific quantifiable Open Science practices or outputs substitute previous misused metrics, missing the overall need to monitor a comprehensive transformation of a new research culture. In addition, there is a need for assessing the values and impacts of science focusing on the people who are doing, engaging with, and/or benefiting from science, and to address the current lack of policy and training monitoring. Existing methods to assess the adoption of Open Science practices should therefore be strengthened (UNESCO, 2023), particularly to track the research culture change and value an open and reproducible research process.

Possible shortcomings of frameworks for Researcher Profiles may lie in the lack of consensus on a unified approach for evaluation, on the limited quality of the metadata aggregated by a wide range of data sources or on the identification of trusted, and non-commercial datasources. To overcome these challenges, we are actively participating in the CoARA Working Group (WG) "Towards Open Infrastructures for Responsible Research Assessment", which involves a comprehensive redefinition and evaluation of the necessary infrastructural components for fair and inclusive research assessment. The WG focuses on setting the foundational principles and defining the breadth and variety of critical components that should be incorporated into an Open Infrastructure fit for Research Assessment. These discussions are of the utmost importance in order to establish a robust framework that promotes transparent, collaborative and equitable research evaluation practices. While Narrative CVs can be a powerful tool to recognise the variety of research activities and outputs, and to reflect and support the diversity of individuals and ideas that contribute to high quality research (Aubert Bonn et al., 2024), diverse organisations may have different understandings about the potential value that using narrative CVs might add to their assessment processes. This requires adopting a sufficiently flexible definition of narrative CVs which can be adapted to various local contexts and cater to funders' diverse needs and goals (Fritch et al., 2021). Furthermore, as the consensus on narrative CVs and qualitative assessments gains traction, it is crucial for research organisations to engage with their research communities to promote the ongoing efforts and transformations. With new research evaluation policies and practices progressively becoming the norm, constant dialogue and critical analysis will be essential to guide the stakeholders to a responsible and ethical use of research assessment. This interactive session will therefore aim at gathering feedback from participants to feed the design of a framework which facilitates research assessment. Conclusions Our participation in the REvaluation conference aims to engage the research and experts community members in shaping the GraspOS Researcher Profile as a tool to support a research assessment system that supports Open Science. We aim to provide a paradigm for research assessments that prioritise diversity, inclusivity, and a comprehensive understanding of research impact. This work is a significant and direct contribution to the conference discussions on advancing research evaluation systems.

REFERENCES:

- European Commission, Directorate-General for Research and Innovation, European Research Area policy agenda – Overview of actions for the period 2022-2024, Publications Office of the European Union, 2022. https://data.europa.eu/doi/10.2777/52110
- Institut de France, Académie des Sciences, "On the proper Use of Bibliometrics to Evaluate Individual Researchers", Report presented to the French Minister for Higher Education and Research, 17 January 2011. https://www.academie-sciences.fr/archivage_site/activite/rapport/avis170111gb.pdf
- Hicks, D., Wouters, P., Waltman, L. et al., Bibliometrics: The Leiden Manifesto for research metrics. Nature 520, 429–431, 2015. https://doi.org/10.1038/520429a
- Pontika, N., Klebel,T., Correia, A., Metzler, H., Knoth, P., Ross-Hellauer, T., Indicators of research quality, quantity, openness, and responsibility in institutional review, promotion, and tenure policies across seven countries. Quantitative Science Studies; 3 (4): 888–911, 2022. https://doi.org/10.1162/qss_a_00224
- Declaration on Research Assessment (DORA), Guidance on the responsible use of quantitative indicators in research assessment, 2024. http://doi.org/10.5281/zenodo.10979644
- Edwards, M. A., & Roy, S., Academic Research in the 21st Century : Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hypercompetition. Environmental Engineering Science, 34(1), 51-61, 2017. https://doi.org/10.1089/ees.2016.0223
- Di Donato F., "What we talk about when we talk about research quality. A discussion on responsible research assessment and Open Science", Bollettino telematico di filosofia politica, March 2024. https://commentbfp.sp.unipi.it/quality-fdd/; https://doi.org/10.5281/zenodo.10890788
- Biagioli, M., & Lippman, A., Gaming the Metrics : Misconduct and Manipulation in Academic Research, The MIT Press, 2020. https://doi.org/10.7551/mitpress/11087.001.0001
- European Commission, Directorate-General for Research and Innovation, Towards a reform of the research assessment system – Scoping report, Publications Office, 2021, https://data.europa.eu/doi/10.2777/707440
- Coalition for Advancing Research Assessment (CoARA), Agreement on Reforming Research Assessment, 2022, https://coara.eu/agreement/the-agreement-full-text 11.

- European Commission, Action Plan by the Commission to implement the ten commitments of the Agreement on Reforming Research Assessment (ARRA), 2024, https://research-and-innovation.ec.europa.eu/document/download/e69aff11-4494-4e5f-866c-6945 39a3ea26_en?filename=ec_rtd_commitments-reform-research-assessment.pdf
- International Network of Research Management Societies - Research Evaluation Group, The SCOPE Framework, The University of Melbourne, Report, 2023. https://doi.org/10.26188/21919527.v1
- Ibid.
- O'Neill, G., OPUS Deliverable 3.1: Indicators and Metrics to Test in the Pilots. Zenodo, 2024, https://doi.org/10.5281/zenodo.10497434 15.
- UNESCO, Recommendation on Open Science, 2021. https://doi.org/10.54677/MNMH8546
- UNESCO, Open science outlook 1 : Status and trends around the world—UNESCO Digital Library, 2023. https://doi.org/10.54677/GIIC6829
- Aubert Bonn N., Morris JP., Sapcariu S., & Stroobants K., Are Narrative CVs contributing towards shifting research culture? Workshop Report from the 2023 Recognition and Rewards Festival [version 1; peer review: awaiting peer review]. F1000Research 2024, 13:332, https://doi.org/10.12688/f1000research.146108.1
- Fritch, R., Hatch, A., Hazlett, H., & Vinkenburg, C., Using Narrative CVs: Process Optimization and bias mitigation, Zenodo, 2021. https://doi.org/10.5281/zenodo.5799414

LluÌs Rovira, Ferran Lavall
AGAUR (Research funding Agency of Catalonia) , Spain

A PROVED METHODOLOGY FOR THE PERFORMANCE ASSESSMENT OF RESEARCH INSTITUTES

ABSTRACT

In 2012 a methodology for assessment and evaluation of research institutes was developed to be implemented in the CERCA system of research centres www.cerca.cat . The methodology of evaluation has been essentially conserved in the last 12 years only incorporating minor modifications.

This methodology is based on the work of an assessment committee that acts collectively and by consensus. Its members have the responsibility to read all the submitted documentation, receive instructions from the members of the centre's governing body with respect to the assessment, interact with the centre's directors and staff during the on-site visit, discuss the different assessable aspects and issue recommendations to be fulfilled in the next period. The experts of this panel are selected among prestigious researchers and professionals all over the world. As an example, the assessors have belonged to a wide range of institutions, including the National Institute for Agricultural Research (INRA), the National Institute for Health and Medical Research (INSERM) and the National Centre for Scientific Research (CNRS), in France; the Max Planck Society, Huawei, the Fraunhofer Institute and the German Archaeological Institute (DAI), in Germany; European institutions such as the European Molecular Biology Laboratory (EMBL), the European Organization for Nuclear Research (CERN), the European Space Agency (ESA) and Fusion for Energy (F4E); Tel Aviv University; Telecom Italia; the ALBA Synchrotron; the Spanish National Cancer Research Center (CNIO); Tecnalia; the Swiss Federal Institute of Technology in Lausanne (EPFL); the Roslin Institute; Imperial College London; the University of Oxford; New York's Memorial Sloan Kettering Cancer Center; the Bill & Melinda Gates Foundation; the Massachusetts Institute of Technology (MIT); Harvard University; Stanford University; Princeton University, the University of California, among many others.

The awarding of the centre's final mark falls to the assessment committee chair, after having heard all the committee members. The committee must comprise, at least, the following members:

1. Up to two scientists (one of them appointed chair).
2. One members of the centre's scientific advisory board or its equivalent
3. One KTT assessor
4. One Management assessor.
5. One non-voting rapporteur

They all sign a declaration of no conflict of interest and confidentiality. The identity of the assessment committee members may not be revealed until the day of the on-site visit. The list of assessment committee members must appear in the final assessment report.

The assessment's primary aim has been to measure how well the centres' have achieved their mission objectives in the previous four years and, if necessary, update them through analysing several issues such as the general concept of the Institute, the results, services, projects, knowledge and technology transfer, intellectual property protection, prizes and positions of importance, and management including some open science issues. Also the implementation of the recommendations from the previous assessment is given a high importance.

The centres under assessment had to fill in a self-assessment report as well as other appendices such as the reports issued by the scientific advisory board (SAB) for the previous four years, as well as any other assessments carried out by other third-party institutions (e.g., SeO (Severo Ochoa) or ISCIII) in Spain.

The evaluation runs through a site visit of one day that takes place in the evaluated Institute or Centre.

When the assessment committee has finished drafting the assessment report, it must send it to the director of the centre under assessment, who then has a maximum seven-day appeal period to point out any errors or inaccuracies that need to be rectified.

After this time, the report must be updated to reflect any new changes, following consultation with the chair. When this is done, the report is considered complete. The final report must be sent to all the members of the highest governing body of the Centre and to the director of the centre. The marks can be the following:

(A) Outstanding performance, placing the centre among the world's most influential and top performing institutions in its particular field in terms of originality, rigour and significance.

(B) Excellent performance, whereby results are produced that nearly meet the highest international standards in terms of originality, rigour and significance.

(C) Very good performance, meaning that it is beginning to garner international recognition, although focus on some strategic issues is required to achieve greater originality, rigour and significance.

(D) Clear need for improvement. The centre should update its focus or make changes, as its current structure and/or performance does not provide guarantees for the Board of Trustees. Every letter mark can be adjusted up or down slightly with a plus (+) or minus (-) at the discretion of the chair. The marking scale is more precise this way, thus reflecting the reality of the centre under assessment with greater accuracy.

There are aprox. 40 CERCA centres. Each centre or institute is evaluated every 4 years. Up to date about 120 evaluations have been implemented and each CERCA centres has been evaluated three times. A strong pattern of this methodology is the assumption that what is required in the form of evaluation becomes important or relevant for the research institutes successful performance. Accordingly several messages have been given to the CERCA institutes through the evaluation. Examples of this are:

- Bibliometrics is used only to measure scientific leadership, gender share in authorship and international collaboration
- Human resources strategy for researchers needs to be awarded to CERCA institutes
- Mobility of young researchers (postdocs) is crucial
- The coordination of EU projects and ERC grants are a measure of success
- Quality not quantity in scientific production. Only 10 papers (link to full text) are considered in the evaluation of the institute in the period of 4 years.
- Only 5 research projects are considered
- Only 5 KTT actions are considered. IPR policy is required as well as Spinouts creation and management.
- Only 5 Outreach actions are considered
- Only 3 EDI actions are considered
- Philanthropic funding must be strategically planned
- Ethical/integrity control is necessary • Institutional cooperation and networking is welcome

This methodology has been COARA respectful during all this years. After 12 years of evaluation the CERCA centres have reached important milestones such as an outstanding funding level at EU-level

Kristina Masevičiūtė
Research Council of Lithuania, Lithuania

THE EVOLUTION AND FUTURE DIRECTIONS OF LITHUANIA'S RESEARCH PERFORMANCE ASSESSMENT SYSTEM: A CASE STUDY

ABSTRACT

In Lithuania, research performance is primarily driven by the higher education (HE) and government (GOVERN) sectors, which account for the highest share of researchers and R&D spending. In 2022, these sectors employed 68% of full-time equivalent (FTE) researchers and represented 52% of R&D expenditure. 80 % of R&D funding in HE and GOVERN sectors are from state budget funds and nearly half of the R&D spending in the HE and GOVERN sectors is derived from direct state budget appropriations. Consequently, the efficiency of public funding distribution and accountability for public investment in research through higher education institutions are crucial factors in enhancing the performance-based research funding system (PFRS) for allocating institutional funding.

Lithuania has introduced the allocation of institutional funding based on ex post assessment of research performance of public research performing organisations (universities and research institutes) since 2005. Since then, the research assessment system has undergone several rounds of modifications, including use of national peers, as well as expanding the spectrum of indicators. A significant change was introduced in 2018, when, in addition to the existing annual assessments of R&D activities, a comparative expert assessment (CEA) was implemented. Therefore, the current Lithuanian assessment system for R&D activities consist of formal, based on quantitative indicators assessment of R&D activities conducted annually and the CEA, performed every five years. The CEA results implemented in 2018 determined 60% of state basic budget funding for R&D activities for the subsequent five years (70% in 2023), with 40% coming from the annual assessments (since 2023 - 30%).

The CEA of R&D activities carried out by Lithuanian universities and research institutes was designed to overcome several key obstacles in the existing research performance assessment system. Firstly, although the weight of research quality criterion is the most significant, the range of assessment criteria was extended and the societal relevance and economic impacts as well as the environment of the assessed unit were included into evaluation. Each participating institution or its part which is formed for assessment purposes (so called unit of assessment) is assessed against the three criteria:

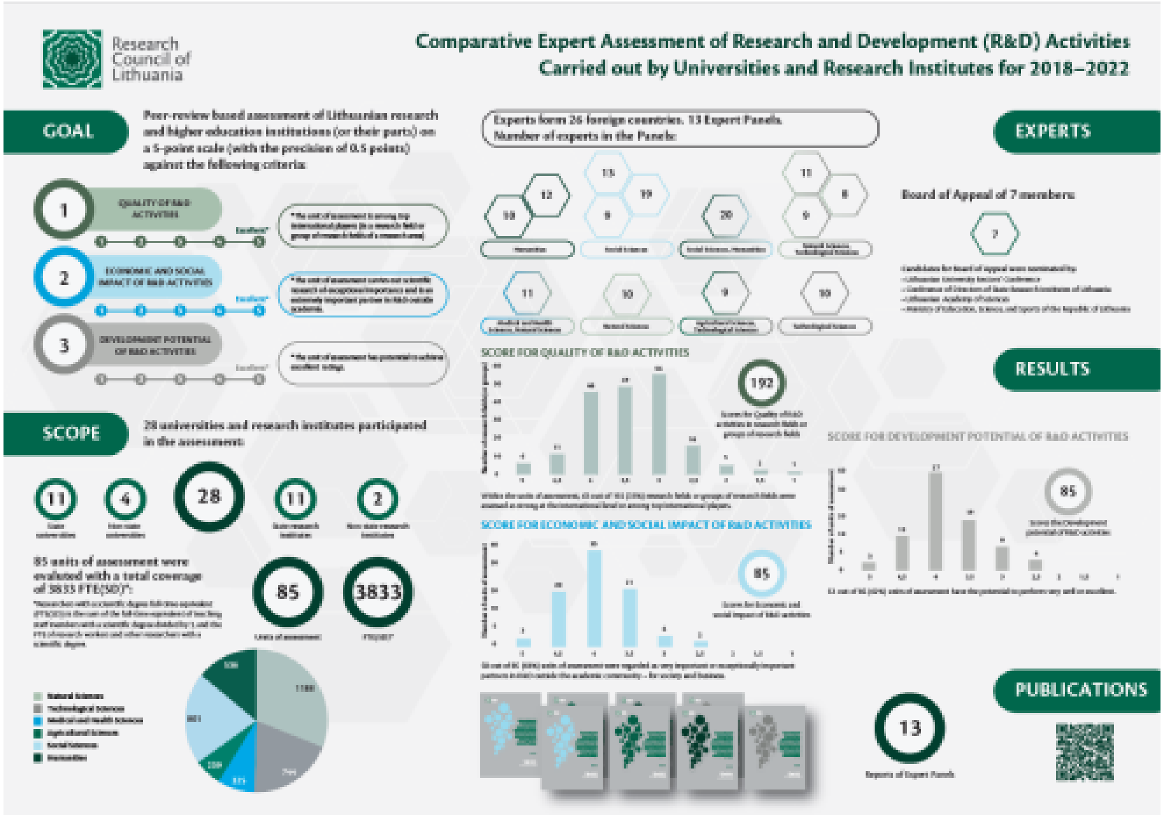
- The quality of R&D activities (weight 0.65);
- The economic and social impact of R&D activities (weight 0.2);
- The development potential of R&D activities (weight 0.15).

To address the challenge of using national peers, which could compromise objectivity due to the country's small size, the assessment included international experts to enhance impartiality. Experts are grouped into Panels which creates a possibility for discussion and debate within the peer group and enabling comparison within the group. In 2023 113 experts were invited to assess the research performance of 85 units of assessment and the experts were grouped into 13 Panels.

Another novelty in the Lithuanian research performance assessment system was introduction of site visits to the Institutions. The visits enabled experts to gain a more comprehensive understanding of the research environments. Additionally, the opportunity to meet with administration, researchers, and PhD students provided valuable supplementary evidence for the assessment, complementing the documentation and written information provided by the unit. Although site visits to institutions in the Lithuanian research and higher education system are used for external evaluations of higher education institutions and study programs, a particular focus on research activities has contributed to capacity building to develop institutional or departmental research strategies. Furthermore, the experts provide not only scores and reasoned justification for each criterion, but also written feedback to those evaluated, helping them identify areas for improvement. According to the feedback survey of units which have participated in 2023 CEA (with participation rate in the survey of nearly 90%), 85 % of units strongly agreed or agreed with the statement "The recommendations made by the expert are valuable for improving the performance of the unit". The scores for each criterion and their designated weights, along with the FTE of researchers holding scientific degrees, are used in the funding formula. The reports from each of the Panel provide valuable evidence to R&D policymakers at various levels.

The information and the results of the last round of CEA could be found at: https://lmt.lrv.lt/en/science-quality/comparative-expert-assessment-of-research-and-developmentactivities/

Infographic about CEA in 2023:



Future directions

Considering the commitments within the Agreement of Reforming Research Assessment, one of the alternatives for modification of Lithuania research assessment system under consideration is to discontinue the annual assessment and rely solely on the results of expert assessment. The pros and cons need to be carefully considered. Additionally, the data provided for expert assessment is currently under review. The report submitted by each unit contains substantial factual data, but the self-reflection elements could be enhanced. While experts consider the quantitative data on overall performance important, achieving a balance between reflective aspects and supporting quantitative evidence is necessary. The assessment of the impact criterion and the data supporting it also require reconsideration.

Dr. Giovanna Lima, Latiffah Salima Baldeh, Balint Hardy
Erasmus University Rotterdam, Netherlands

ENABLING IMPACTFUL ACADEMIC CAREERS

ABSTRACT

The academic incentives and rewards system is often narrow in its focus on research productivity, its application of a limited number of standardised metrics and its summative approach aimed at selection. The assessment of academics should recognise the diverse outputs, practices and activities that maximise the quality and impact of scholarly contributions (CoARA, 2023; Kramer, B. and Bosman, J., 2024). This is particularly relevant for an increasing relevant priority within universities that encourages academics to address societal challenges in their research, teaching, and engagement.

Choices in assessment of academic activities (whom to assess, what to assess and how to assess) have the potential to shape both the institution and the wider system of higher education (Kramer and Bosman, 2024). Review, promotion, and tenure (RPT) processes significantly affect how faculty direct their own career and scholarly progression (Schimanski & Alperin, 2018). If universities are indeed committed to supporting, facilitating, and enabling positive societal impact, it is time to learn, develop, and test new methods and approaches. We need to facilitate, appreciate, recognise, and reward impactful careers.

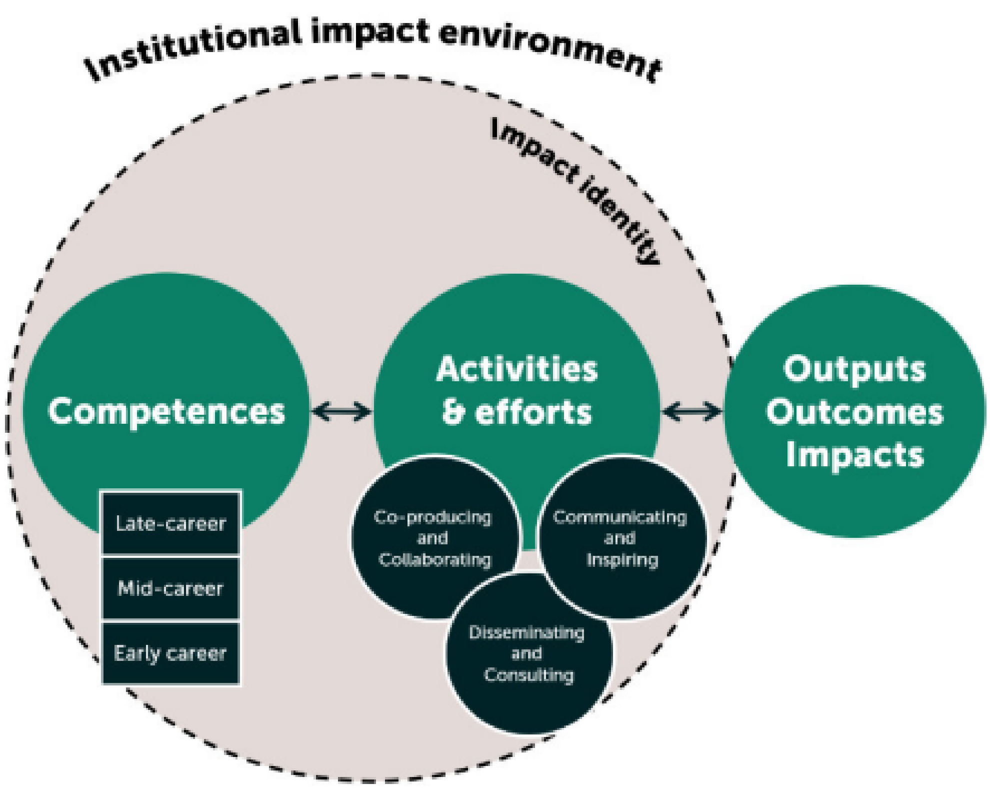
Organisations have an important role and the power to shape impact practices of its members (de Jong & Balaban, 2022) by providing enabling conditions (practical support and resources) that facilitate the emergence of productive interactions; the exchange between researchers and stakeholders where socially relevant knowledge is produced (de Jong et al., 2022; Spaapen & van Drooge, 2011). As with multiple universities across the world, Erasmus University Rotterdam (EUR) aims to transition to an impact-driven university over the next few years. To accomplish this objective, we recognise the need to revise the current system for recognising and rewarding our academics' impact-related competences, talents, activities, and outputs, to ensure their efforts and ambitions are institutionally supported and enabled. The EUR Framework on Recognition & Rewards and the definition of societal impact at EUR are fundamental steppingstones in the initiation of this fundamental change in our HR-policies (Erasmus University Rotterdam, 2021, 2023). These documents lay out the principles, priorities, and essential prerequisites for the required cultural and environmental changes.

To further the capacity of EUR to support and enable creating positive societal impact through our research, education, and engagement, the teams at Recognition & Rewards (R&R) and Evaluating Societal Impact (ESI) developed an internal working framework. It identifies a framework of essential

competences; activities; and outputs and outcomes that can be developed and delivered, and that can be recognised, appreciated, and rewarded by staff and Schools. It aims to highlight how individual impact journeys may differ from one another. It also helps understand that multiple impact identities can flourish, particularly once impact is understood as a team effort. We believe participants from the REvaluation Conference could benefit from our approach but also help us improve it.

The Impactful Careers Framework is intended for multiple audiences. For those in leadership and policy roles, it hopes to inspire them to consider how these topics can be adopted in their policies and procedures. For academic planning their own (or their supervisees) personal development and thinking of their career prospects, the framework enables them to think how they would like to be an impactful academic. We hope to inspire and help Schools, Departments, teams, and individuals in facilitating, acknowledging, recognising, and rewarding impactful academic careers at EUR and beyond. At the moment of submission, the Framework has been used to develop workshops on working in teams and on developing differentiated career paths for academics. So, it has not yet been presented in full to our community, but rather as subsidy to designing policies.

FIGURE 1 - BUILDING BLOCKS OF IMPACT IDENTITIES FOR ACADEMICS



Societal impact considerations can inform recruitment, selection, and promotion procedures, and development appraisal cycles, shaping individual career advancement options. For each of these processes, there are different aspects of impactful careers one can decide to invest on and assess. These include, at least, the competences academics (want/need to) develop when pursuing impactful careers; the activities or efforts they engage and invest in and outputs they produce; and the outcomes and impacts of employing those competences and activities (see Figure I). These aspects can be applied differently depending on the processes we are incorporating them to. For example, if an HR business partner wants to develop guidelines to understand (or assess) competences development and attainment on the context of the development cycle or on a promotion cycle. Or if a Department Head wants to invest in support and monitoring for impact activities.

The contribution to the community of the Impactful Careers Framework is to make explicit "impact competences" and "impact activities" that are yet to be incorporated in the development and assessment of academic careers in an easy to apply tool. We recognise societal impact can be created through a lot of ways and indeed most (scholarly) competences and activities have the potential to contribute to advancing our societal impact either on the short or long term. What we call in the framework "impact competences" and "impact activities" aims to recognise there are a particular set of activities that are especially useful in realising this potential and are embedded in and strengthen impact pathways.

On Impact Competences. As defined by the Competence Instrument for the Dutch Universities, a competence is a unique combination of knowledge, skills, and attitude, which find their expression in outward behaviour and are key to a person's success in their job and career. It is increasingly common to see frameworks being developed for (researchers') competences, including but not limited to ResearchComp: The European Competence Framework for Researchers, 'Science4Policy' Competence Framework, GreenComp The European sustainability competence framework, and EntreComp: The entrepreneurship competence framework. In our framework, we compiled these and additional frameworks on open and responsible science, inter- and transdisciplinarity to provide an overview of the competences that are essential for impact-driven academic activities. Next to each competence and their definition is their relevance to societal impact and indicators on how such competence could be assessed. We recognise all competences listed can be developed at all career stages, especially as they are related to people's talents, ambitions, and experience.

On impact activities. There are many terms that refer to what we call impact activities, including public engagement, valorisation, science communication, knowledge mobilisation and utilisation, outreach, and so on. They mostly refer to activities which include elements of openness and societal engagement. In its essence, impact activities for our purposes are those activities performed by university staff and that the knowledge and expertise is based on academic and scientific processes, even as the activities themselves are also performed by other organisations such as consultancy agencies or NGOs. In our framework, we have listed common impact activities that can be acknowledged, recognised, and rewarded. We grouped them into broad categories based on the type of (deliberate) engagement with society and their (intended) effects. The activities are common to the three core portfolios of universities (education, research and engagement) and are commonly used, including in funding proposals and training about planning, monitoring, and communicating societal impact.
We see a fit to Strand 2 calls for evaluating careers beyond one-sided frames of reference, that recognises the multidimensional nature of academic careers and their impact. We would be honoured to share the framework in the REvaluation Conference and are quite interested in the exchange of knowledge and experience with other practitioners in the field. For this reason, alongside (or instead of) an interactive presentation, a workshop session where we get to examine the different components of the framework could also be a fruitful option. We also welcome suggestions from this peer review process of what would be the most valuable approach, as we are keen to hear what would be most valuable for the community.

REFERENCES:

- CoARA (2023). Agreement on Reforming Research Assessment. https://coara.eu/app/uploads/2022/09/2022_07_19_rra_agreement_final.pdf
- de Jong, S. P. L., Balaban, C., & Nedeva, M. (2022). From ‘productive interactions’ to ‘enabling conditions’: The role of organizations in generating societal impact of academic research. Science and Public Policy, 49(4), 643–645. https://doi.org/10.1093/SCIPOL/SCAC015
- de Jong, S. P. L., & Balaban, C. (2022). How universities influence societal impact practices: Academics’ sense-making of organizational impact strategies. Science and Public Policy, 49(4), 609– 620. https://doi.org/10.1093/SCIPOL/SCAC012
- Erasmus University Rotterdam (2021). Impact through Recognition & Rewards – EUR Framework, https://www.eur.nl/media/98856

- Erasmus University Rotterdam (2023). Defining Societal Impact at EUR – A common framework for our impact strategy, https://www.eur.nl/en/about-eur/strategy-2024/about-strategy-2024/definingsocietal-impact-eur
- Grant, J. (2012). “Academic Incentives and Research Impact: Developing Reward and Recognition Systems to Better People’s Lives,”. Academy Health, February 2021.
- Kramer, B. and Bosman, J. (2024). "Recognition and Rewards in Academia – Recent Trends in Assessment", Thunnissen, M. and Boselie, P. (Ed.) Talent Management in Higher Education (Talent Management), Emerald Publishing Limited, Leeds, pp. 55-75. https://doi.org/10.1108/978-1-80262-685- 820241004
- Schimanski, L. A., & Alperin, J. P. (2018). The evaluation of scholarship in academic promotion and tenure processes: Past, present, and future. In F1000Research (Vol. 7, p. 1605). F1000 Research Ltd. https://doi.org/10.12688/f1000research.16493.1
- VSNU-NOA (2015). Competence Instrument for the Dutch Universities. https://www.universiteitenvannederland.nl/files/publications/Vernieuwd%20Competence%20Instrume nt%20Dutch%20Universities%202016-%20EN.pdf

Florian Bayer
University of Vienna, Department of Science & Technology Studies , Austria

VALUING RESEARCH IN SSH FIELDS: HOW SSH SCHOLARS VALUE RESEARCH AND PUBLICATION PRACTICES IN THE FACE OF EPISTEMIC CAPITALISM

ABSTRACT

This proposal for a contribution is based on my PhD project in science & technology studies (STS) investigating indicator use in everyday (research) practice in the social sciences and humanities (SSH). Extending existing lines of STS work on indicators and metrics (de Rijcke et al., 2016) to SSH fields based on case studies in history, political science and area studies at the University of Vienna, I am asking how indicators, metrics and other forms of quantification matter at different scales: What role do different sociotechnical practices of quantification play in everyday research practice at the level of individual SSH researchers? How are practices in research shaped by institutional contexts and settings? How is the quantification of research institutionalized and practiced in context of the national research landscape? By addressing these questions I want to trace the presence and role of indicators and metrics in SSH research, in order to identify different forms and processes of quantification of the epistemic living spaces (Felt, 2009) in question. Against this background, the suggested contribution is situated at the level of individual researchers, and focuses on the ways SSH scholars are making sense of their own research and its outcomes in everyday practice.

The presentation of results will be preliminary, as analysis will not be fully completed by December 2024. So far a total of 46 qualitative in-depth interviews have been conducted with 44 different researchers (23 male, 21 female) between September 2018 and December 2023. Beginning with historians in the first wave (September 2018 – April 2019), I moved on to interview researchers across the three case studies in a second (November 2019 – February 2020) and a third wave (January 2023 – December 2023). For each case I approached a number of senior scholars to explore the characteristics of the research field, its institutionalization at the University of Vienna and to get their perspective as researchers, who have been managing and shaping departments, who have experience in different sorts of hiring processes, as well as in leading and mentoring PhDs and postdocs in their fields. The data set comprises material of 14 interviews with six historians, with four political scientists and three senior scholars in area studies. The rest of the interviews targeted researchers at different career stages beginning with late PhDs, who have had made first experiences with regard to academic publishing, researchers on postdoc positions as well as more advanced scholars on tenure track positions. In all of these

categories I made sure to include perspectives of male and female researchers, of scholars on university assistant positions and of staff working in third-party funded projects alike. As a result, the data set comprises ten such interviews with historians, 13 interviews with political scientists and material from ten interviews with nine scholars in area studies.
Interviewees gave written consent on their participation. Conversations were recorded with digital recording devices. The interviews – ranging from one hour up to two and a half hours – were consequently transcribed and imported to Atlas.ti for analysis. By the end of May 2024 the recordings of all of the interviews have been transcribed, while 33 transcripts have been proof-read. Out of the proof-read transcripts 22 have also been coded in Atlas.ti. Coding was started while I was still gathering material and conducted according to the principles of Grounded Theory (Strauss & Corbin, 1998; Bryant & Charmaz, 2010). The intention was to structure the gathered data based on open, inductive and deductive coding right away, so that emerging categories and themes could inform ongoing sampling and data collection based on a zig-zag approach (Rivas, 2018). Throughout the first rounds of data collection and analysis in the fields of history and political science, the coding process was restarted and restructured several times. These repeated iterations and adaptations were developed by moving from traditional Grounded Theory towards Abductive Analysis, which offered a good way into structuring the material based on the process of alternative casing (Tavory & Timmermans, 2014, S. 58–61).
To keep the interviews focused on first hand experiences and practices, I developed an interview guideline accompanied with a set of cards. These cards were used to actively confront the interviewees with quotations and insights from earlier interviews along the emerging categories and themes of the coding process. Card-based methods have been successfully facilitated to render nondebatable issues debatable in focus group discussions on nano-technologies (Felt et al., 2018), and to enable researchers to reflexively discuss matters of responsible research and innovation (Felt et al., 2018) and research integrity (Felt & Frantz, 2022). The interview cards were utilized throughout the conversations to jump-start reflections and to offer orientation on what aspects and levels of abstraction to focus on, enabling the interviewees to relate and position their own sense-making to that of others.
Looking into the ways in which SSH researchers reflected on research and publication practices when reporting about their personal experiences throughout their careers – e.g. in applying for jobs, the ways they had (not) planned or prepared for certain career steps and the role they attribute to different research and publication practices in their everyday working routines – I turned to Heuts and Mol's (2013) notion of valuation as a practice in

order to further sensitize quality judgments as a central aspect of research practice. Conceptualizing valuation as a practice the authors are moving away from looking into certain qualities of things and towards "foregrounding 'valuing'" as an activity by identifying and closely looking into different "registers of valuing". These registers "indicate a shared relevance, while what is or isn't good in relation to this relevance may differ from one situation to another" (Heuts & Mol, 2013, S. 129). Speaking of "valuing" Heuts and Mol highlight how assessment, judgment, valuation, evaluation, improvement and the like "slide over into each other" in practice (p. 130). This way different registers – drawn upon to render things as relevant, specifying the specific ways that make them good or not good – can be disentangled by analysis, and reflected with regard to overlaps and internal tensions (p. 129).

Based on the approach developed by Heuts & Mol the narrations and reflections of researchers about their own research and publication practices are investigated as moments and acts of valuation and as accounts thereof. In doing so we can analyze the different registers that allow scholars to value certain aspects of research as good and bad or of high and low quality. I speak of research and publication practices, because my empirical research aimed to address these questions from the perspective of everyday research practice. Rather than studying moments of assessment – e.g. in hiring committees or on funding panels – I address these questions from the perspective of knowledge production. The interview material gathered in the three case studies covers different sorts of practices in SSH research, e.g. "searching and finding literature", "reading practices", "empirical research" and "publication practices".

The results presented and discussed will for the most part be based on the analysis of material concerned with publication practices. Also in this regard the aim was to approach the question of quality and valuation from the perspective of the production process. Instead of studying moments of assessment or asking researchers what constitutes a good monograph or journal article, scholars were encouraged to recall the practical histories of their own publications: e.g. starting with the initial broader context of the respective research, to the research process to the first idea for a manuscript to the writing, review and production process. Analyzing these accounts, all sorts of ways in which SSH scholars are valuing practices related to publishing could be uncovered. Throughout the coding process three registers of valuing publication practices emerged: an epistemic register, a reputational register, and an institutional register. The epistemic register is concerned with the processes of knowledge production, the form and nature of these processes and the knowledge produced; the reputational register is concerned with gaining and loosing status within knowledge communities and in academia in general; the institutional register is used to denote meaning and relevance with regard to institutional contexts, settings and practices.

Heuts and Mol emphasize how describing and identifying different registers in empirically studying valuations serves analytical purposes and does not imply that these registers are present in pure and separate ways in actual social practice. Quite on the contrary, actual valuation practices are messy and complex and the registers of valuing drawn upon are overlapping or also in tension with each other (Heuts & Mol, 2013, S. 129). In a similar vein Helgesson (2016) has suggested to investigate the ways in which multiple valuations and different valuation practices are folded into each other, because "looking into the nooks and crannies of a conglomeration of interrelated valuation practices further provides a glimpse of a politics beyond the singular valuation practice; this is the politics of how valuation practices are folded on to one another, and how these folds are characterized" (p. 100–101). Other scholars highlighted that we do not only have to take into account the multiplicity of valuing as a practice as such, but also account for the multiple contexts in which these valuations take place. Denoting the concept of valuation constellations Waibel and colleagues urge us to take into account the positions and relations between the valuee, the valuator and the audience in analyzing valuations, and to include the role of valuation rules and valuation infrastructures in the analysis (Waibel et al., 2021). Consequently, the presentation of how SSH scholars value research and publication practices will conclude by analyzing certain forms of overlap between the registers of valuing publication practices identified, specific foldings of different kinds of valuations, and their relation to imagined, anticipated and actually experienced valuation constellations.

The further analysis of the empirical findings will follow STS scholarship using the notion of capitalism as a sensitizing concept. Building on early work in laboratory studies and the notion of the cycle of credit (Latour & Woolgar, 1986) as well as Hackett's notion of academic capitalism (1990, 2014) Fochler (2016) denoted the notion of epistemic capitalism, that allows us to trace how contemporary research has come to resemble certain forms and dynamics of the capitalist mode of production. Shifting the focus from earlier studies investigating the entanglement of academia and corporate capital (Slaughter & Leslie, 1997; Slaughter & Rhoades, 2004), the inquiry into epistemic capitalism is concerned with the ways in which researchers navigate different value orders in everyday practice, as they strive to accumulate epistemic capital understood as "worth made durable, through the act of doing research" (Fochler, 2016, S. 924). Following such a conceptualization of researchers as "entrepreneurial managers of their own careers, publications, and grant portfolios" (ibid.) Rushforth, Franssen & de Rijcke emphasize how researchers may consider various kinds of worth to be important, while "not all forms can be accumulated and reinvested in subsequent cycles of credit" (Rushforth et al., 2019, S. 212). Investigating biomedical research groups they analyze how laboratories – conceived as "configurations of materials, techniques, instruments and ideas and enabling theories drawn upon in producing new research" – make possible and limit efforts of research portfolio construction (p. 213). In doing so the notion of the portfolio and related portfolio strategies prove fruitful to sensitize the ways in which researchers address multiple and at times conflicting orders of worth.

Contributing to this line of work I will discuss how SSH scholars mobilize, enroll and address different forms of research quality in everyday research practice. In contrast to laboratories, the notion of the career trajectory serves as the central reference point that enables, structures and limits SSH scholars' portfolio strategies. Recently, scholarship on the impact of quantification on the social sciences in the United Kingdom (Pardo-Guerra, 2022) as well as research on strategic decision making in relation to the academic labor market in general (Gl‰ser & Laudel, 2015; Laudel & Bielick, 2018, 2019) has convincingly utilized a concept of academic careers based on the works of the Chicago School. Transcending earlier narrow conceptualizations of careers as a sequence of jobs or professional statuses, by more broadly and firmly grounding the concept in various social settings, authors like Hughes and Goffmann had stressed the capacity of this notion to conceptualize the nexus between the personal and the collective, individual action and social structure (Hughes cf. Barley, 1989, S. 46; Goffman, 1961, S. 127–128). Understood this way, the notion of academic careers enables us to conceptualize how SSH scholars negotiate the nexus between individual and collective research practices, between the academic self and the community it is embedded in, as well as between individual action and structural requirements and expectations. Analyzing how SSH scholars value research and publication practices in relation to academic career trajectories, I will carve out how they develop different portfolio strategies (Rushforth et al., 2019) to strategically plan and manage their publishing activities in relation to future career trajectories. Time and energy invested in certain research activities, much like spending time on publication activities, demand careful consideration and planning, in order to maximize the potential "return on investment" in terms of a successful career trajectory. This way the act of doing SSH research enables various processes and forms of producing and accumulating epistemic capital (Fochler, 2016). Presenting these findings I will discuss and relate to central questions of the thematic strand: How are research quality, good conduct and success in academia defined, negotiated and reshaped in everyday research contexts and practice? How does the multiplicity of academic careers relate to epistemic practices? What dynamics result from knowledge production taking the form of producing and accumulating epistemic capital?

REFERENCES:

- Barley, S. R. (1989). Careers, identities, and institutions: The legacy of the Chicago School of Sociology. In M. B. Arthur, D. D. Hall, & B. S. Lawrence (Hrsg.), Handbook of Career Theory (S. 41–65). Cambridge University Press.
- Bryant, A., & Charmaz, K. (2010). The SAGE Handbook of Grounded Theory: Paperback Edition. SAGE Publications. de Rijcke, S. de, Wouters, P. F., Rushforth, A. D., Franssen, T. P., & Hammarfelt, B. (2016). Evaluation practices and effects of indicator use—A literature review. Research Evaluation, 25(2), 161–169. https://doi.org/10.1093/reseval/rvv038
- Felt, U. (2009). Knowing and living in academic research. In Knowing and Living in Academic Research. Convergence and Heterogeneity in Research Cultures in the European Context (S. 17–39). Inst. of Sociology of the Acad. of Sciences of the Czech Republic.
- Felt, U., Fochler, M., & Sigl, L. (2018). IMAGINE RRI. A card-based method for reflecting on responsibility in life science research. Journal of Responsible Innovation, 5(2), 201–224. https://doi.org/10.1080/23299460.2018.1457402
- Felt, U., & Frantz, F. (2022). RESPONSE_ABILITY A Card-Based Engagement Method to Support Researchers' Ability to Respond to Integrity Issues. Science and Engineering Ethics, 28(2), 14. https://doi.org/10.1007/s11948-022-00365-6
- Fochler, M. (2016). Variants of Epistemic Capitalism. Science, Technology, & Human Values, 41(5), 922–948. https://doi.org/10.1177/0162243916652224
- Glaser, J., & Laudel, G. (2015). The Three Careers of an Academic. 35, 1–40. http://www.laudel.info/wpcontent/uploads/2015/12/35_2015discussion_paper_Nr_35_Gl aeser_Laudel.pdf
- Goffman, E. (1961). Asylums: Essays on the social situation of mental patient and other inmates. Anchor Books.
- Hackett, E. J. (1990). Science as a Vocation in the 1990s. The Journal of Higher Education, 61(3), 241–279. https://doi.org/10.1080/00221546.1990.11780710
- Hackett, E. J. (2014). Academic Capitalism. Science, Technology, & Human Values, 39(5), 635–638. https://doi.org/10.1177/0162243914540219
- Helgesson, C.-F. (2016). Folded Valuations? Valuation Studies, 4(2), Article 2. https://doi.org/10.3384/VS.2001- 5992.164293

- Heuts, F., & Mol, A. (2013). What Is a Good Tomato? A Case of Valuing in Practice. Valuation Studies, 1(2), 125– 146. https://doi.org/10.3384/vs.2001-5992.1312125
- Latour, B., & Woolgar, S. (1986). Laboratory Life. The construction of Scientific Facts.
- Laudel, G., & Bielick, J. (2018). The Emergence of Individual Research Programs in the Early Career Phase of Academics. Science, Technology, & Human Values, 43(6), 972–1010. https://doi.org/10.1177/0162243918763100
- Laudel, G., & Bielick, J. (2019). How do field-specific research practices affect mobility decisions of early career researchers? Research Policy, 48(9), 103800. https://doi.org/10.1016/j.respol.2019.05.009
- Pardo-Guerra, J. P. (2022). The Quantified Scholar: How Research Evaluations Transformed the British Social Sciences. In The Quantified Scholar. Columbia University Press. https://doi.org/10.7312/pard19780
- Rivas, C. (2018). Finding themes in qualitative data. In C. Seale (Hrsg.), Researching Society and Culture (S. 431–453). SAGE Publications.
- Rushforth, A., Franssen, T., & de Rijcke, S. (2019). Portfolios of Worth: Capitalizing on Basic and Clinical Problems in Biomedical Research Groups. Science, Technology, & Human Values, 44(2), 209–236.
- Slaughter, S., & Leslie, L. L. (1997). Academic Capitalism: Politics, Policies, and the Entrepreneurial University. Johns Hopkins University Press.
- Slaughter, S., & Rhoades, G. (2004). Academic capitalism and the new economy: Markets, state, and higher education. Johns Hopkins University Press.
- Strauss, A., & Corbin, J. (1998). Basics of Qualititaive Research. SAGE Publications.
- Tavory, I., & Timmermans, S. (2014). Abductive Analysis. University of Chicago Press.
- Waibel, D., Peetz, T., & Meier, F. (2021). Valuation Constellations. Valuation Studies, 8(1), 33–66. https://doi.org/10.3384/VS.2001-5992.2021.8.1.33-66

Stefano Bianchini, Valentina Di Girolamo, Julien Ravet, David Arranz
University of Strasbourg, France
European Commission – DG Research and Innovation , Belgium
European Commission – DG Research and Innovation , Belgium
European Commission – DG Research and Innovation , Belgium

ARTIFICIAL INTELLIGENCE IN SCIENCE: PROMISES AND PERILS FOR PRODUCTIVITY

ABSTRACT

Science can help provide us with solutions to global challenges. Yet, empirical evidence suggests that the productivity of scientific research has been decreasing in the last few decades (Pammolli et al., 2011; Boeing and Hünermund; 2020; Aghion et al., 2021), ideas are becoming increasingly difficult to find (Bloom et al., 2020), and the degree of disruptiveness of both scientific papers and patents has decreased significantly over time (Park et al., 2021).

Against this background, the fast-paced proliferation of Artificial Intelligence (AI) and Machine Learning (ML) is set to play a central role in shaping the way research is conducted (OECD, 2023; Arranz et al., 2024). AI represents an extremely powerful and versatile research tool with great potential for discovery, acting as a new "general method of invention" that can influence knowledge creation in many different ways (Cockburn et al., 2018; Bianchini et al., 2022; Krenn et al., 2022).

While a growing body of literature has focused on assessing the degree of penetration of AI technology and the costs and benefits of its diffusion, fewer attempts have been devoted to empirically investigating how AI can actually impact scientific productivity (Bianchini et al., 2022; Yu, 2024). The main objective of this paper is to investigate how the diffusion of AI technologies is affecting scientific production in terms of quantity, quality, and novelty.

In doing so, it contributes to the emerging literature in several ways. First, unlike existing analyses which often have a narrower focus either in terms of the types of AI technologies studied or application domains, our study takes a more holistic approach: we adopt a broader definition of AI and assess its impact across a large number of research fields. Second, we measure research output using a wide range of productivity indicators, encompassing "traditional" citation counts as well as more up-to-date novelty indicators. Third, we examine the impact of AI at three distinct levels: the paper-level (micro), the author-level (meso), and the field-level (macro). Hence, this approach allows us to quantify differences between AI and non-AI papers within the same field, observe changes in research behavior and productivity among individual researchers who adopt AI technologies, and discern macro-level shifts in scientific disciplines as a whole. Indeed, the three levels have different implications for policy. Finally, our work contributes to the

literature by conducting the analysis both at global level and comparing different economic regions, notably the EU, the US and China, thereby providing useful insights also from a policy perspective.

Data and methods: We collected data from OpenAlex (https://openalex.org) and used the hierarchical structure of its "concept" taxonomy – which organizes over 65,000 unique concepts with varying levels of granularity – to identify publications in different scientific domains. Our preliminary analysis focuses on the following arbitrarily chosen domains: Advanced composite materials (1,260 papers); Climate change mitigation (13,438); Cosmology (92,012); Drug discovery (70,645); Hearth diseases (79,517); Numerical weather prediction (10,426); and Particle accelerators (22,878). These domains were chosen to represent both very granular topics, such as advanced composite materials [level-3], and more aggregated topics, such as cosmology [level-1].

We retrieved scholarly publications associated with the above concepts for the period 2000-2022, including peer-reviewed academic journal articles, conference proceedings, and preprint collections. Each publication may be associated with multiple concepts, each accompanied by a score indicating the confidence level for that concept; thus, we could specifically flag papers categorized under [level-1] concepts related to Artificial Intelligence or Machine Learning.

For each paper in our sample, we built a set of metrics to reflect the impact and novelty of the contribution. Impact is measured by the weighted number of citations ('Weighted Nb. Citations') a paper has received from its year of publication up to 2024, the time of data extraction. Also, we created a dummy variable for "big hit" contributions, defined as highly cited papers ('Most Cited'). A paper is considered a "big hit" if it is among the top 10% most cited papers, calculated with reference to other papers published in the same year and within the same concept.

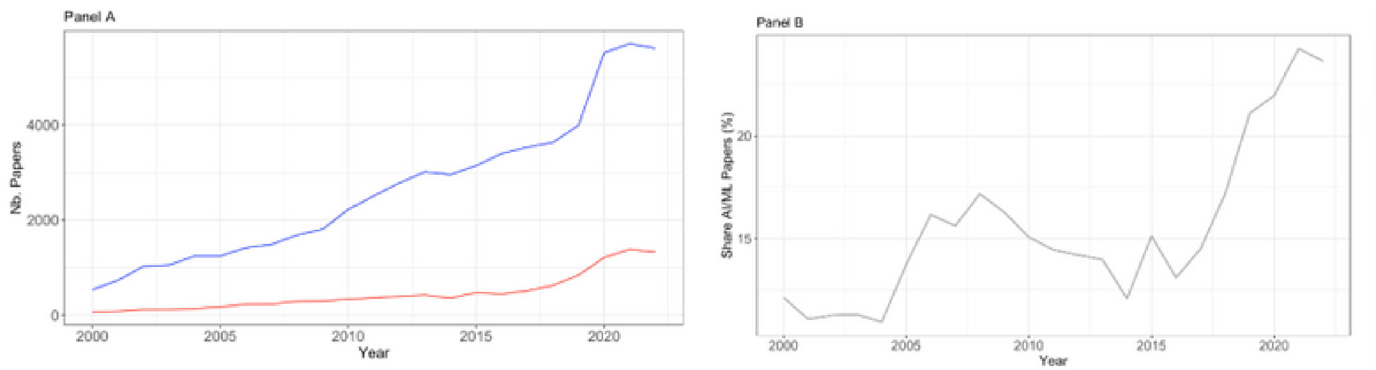
Novelty is operationalized using several indicators. Some are based on atypical combinations of prior knowledge, such as the combination of journals cited in the references ('Novelty Foster' and 'Novelty Lee') (Foster et al., 2015; Lee et al., 2015). Others are based on the appearance of new words or ngrams ('New Words', 'New Bigrams', and 'New Trigrams') and the semantic distance between the focal paper and the prior art that is closest in scientific content ('Cosine Sim. Max') (Art et al., 2023).

The paper-level (micro) compares AI vs. non-AI papers within a given domain using a standard econometric framework. Hence, our main dependent variables are various measures of scientific impact and novelty. The main explanatory variable is a binary indicator that takes the value 1 if the paper involves the use of AI/ML and 0 otherwise. We consider a set of control variables to capture various characteristics of a focal paper, including the team size, the number of references made by a paper, international collaborations, and others.

Preliminary findings: In this section, we report some preliminary results focusing on the domain of “drug discovery”. Results for the other domains broadly conform with the trends discussed here.

As shown in Figure 1, we first observe that the volume of scientific activity related to AI as applied to drug discovery has been growing at a significant pace in recent years, especially starting from the 2010s (Panel A). The share of AI publications in the field reaches 25% in 2022 (Panel B). This is impressive, as it means that 1 out of 4 articles involves the use of AI/ML.

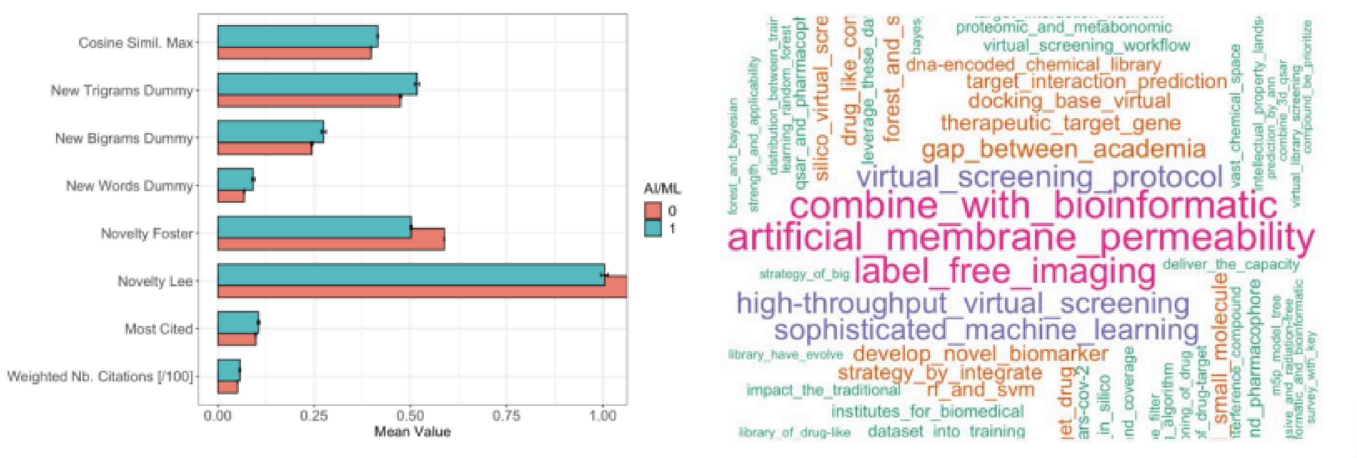
FIGURE 1. DIFFUSION OF AI/ML IN THE DOMAIN “DRUG DISCOVERY”



Notes: Panel A shows the number of papers in drug discovery (blue curve) and those co-classified as AI/ML (red curve). Panel B shows the share of AI/ML papers in drug discovery over time.

In Figure 2 (Panel A), we report the mean difference between AI and non-AI publications with respect to the main variables discussed above. We highlighted three main findings, also confirmed with econometric models (not shown here). First, AI papers receive, on average, more citations and are more likely to be “big hits”, thus attracting greater attention from the scientific community. Second, there is a negative association between AI adoption and atypical combinations of prior knowledge, suggesting that researchers using AI tend to focus on established approaches and proven methodologies. Yet, and third, AI contributions are most distant from the knowledge frontier in the semantic space and are more likely to produce novel outcomes, as measured by new n-grams appearing for the first time in history. This result is particularly interesting because it suggests that despite the tendency to use established methodologies, AI-driven research pushes the boundaries of science in terms of content and innovation (see, e.g., Panel B).

FIGURE 2. COMPARISON OF AI/ML VS. OTHER PAPERS IN THE DOMAIN “DRUG DISCOVERY” AND MOST FREQUENT NEW TRI-GRAMS



Notes: Panel A shows the differences in impact and novelty between AI/ML papers and other papers within the drug discovery domain. Differences are always statistically significant. Panel B shows the most frequently occurring new trigrams (three-word sequences) introduced in these papers.

Research-in-progress – For the author (meso) and field-level (macro) analysis, and to approach as closely as possible the causal effect of AI on scientific research, we plan to leverage two exogenous shocks: (i) the 2012 edition of the ImageNet contest and (ii) the introduction of the first transformer architecture in 2017. In fact, although the scientific community, or at least part of it, had already recognized the potential of AI for research, it could not anticipate the superior performance of the technology following these events. The unanticipated rise of AI provides us with an exogenous event that prompted some scientists to adopt AI for their research, while others with similar characteristics – e.g., publishing in the same journal outlets and having the same seniority – did not. We plan to apply a difference-in-difference (DiD) approach with matching, which should partially balance unobserved confounders and help ensure the parallel trend assumption is met.

rEFERENCES:

- Aghion, P., Céline, A,, and Bunel. S. (2021), The power of creative destruction: economic upheaval and the wealth of nations. Harvard University Press.
- Arranz, D., Bianchini, S., Di Girolamo, V. and Ravet, J. (2023), Trends in the use of AI in science: a bibliometric analysis, European Commission, Directorate-General for Research and Innovation, Publications Office of the European Union, https://data.europa.eu/doi/10.2777/418191
- Arts, S., Melluso, N., & Veugelers, R. (2023). Beyond Citations: Text-Based Metrics for Assessing Novelty and its Impact in Scientific Publications. arXiv preprint arXiv:2309.16437.

- Bianchini, S., Müller, M., & Pelletier, P. (2022), 'Artificial intelligence in science: An emerging general method of invention', Research Policy, 51(10), 104604.
- Bloom, N., Jones, C. I., Van Reenen, J., and Webb, M. (2020). Are ideas getting harder to find?. American Economic Review, 110(4), 1104-1144.
- Boeing, P., and Hünermund, P. (2020), 'A global decline in research productivity? Evidence from China and Germany', Economics Letters, Vol. 197/109646
- Cockburn, I. M., Henderson, R., and Stern, S. (2018). The impact of artificial intelligence on innovation (Vol. 24449). Cambridge, MA, USA: National bureau of economic research.
- Foster, J. G., Rzhetsky, A., & Evans, J. A. (2015). Tradition and innovation in scientists' research strategies. American Sociological Review, 80(5), 875-908.
- Krenn, M., Pollice, R., Guo, S. Y., Aldeghi, M., Cervera-Lierta, A., Friederich, P., ... & Aspuru-Guzik, A. (2022). On scientific understanding with artificial intelligence. Nature Reviews Physics, 4(12), 761-769.
- Lee, Y. N., Walsh, J. P., & Wang, J. (2015). Creativity in scientific teams: Unpacking novelty and impact. Research Policy, 44(3), 684-697.
- OECD (2023), Artificial Intelligence in Science: Challenges, Opportunities and the Future of Research, OECD Publishing, Paris, https://doi.org/10.1787/a8d820bd-en.
- Pammolli, F., Magazzini, L., and Riccaboni, M. (2011). The productivity crisis in pharmaceutical R&D. Nature reviews Drug discovery, 10(6), 428-438.
- Park, M., Leahey, E., and Funk, R. (2021), The decline of disruptive science and technology, arXiv preprint arXiv:2106.11184.
- Yu, Z. (2024). The Impacts of Prediction AI on Scientists: Evidence from AlphaFold. Available at SSRN 4711334.

Stefanie Schuerz, Barbara Kieslinger, Teresa Schaefer
Zentrum für Soziale Innovation GmbH, Austria

IMPACT PATHWAYS ASSESSMENT FOR CITIZEN SCIENCE

ABSTRACT

This submission shares practical insights into the evaluation and impact pathways assessment efforts of the Horizon Europe-funded project ECS – European Citizen Science (2022-2026). The goal of ECS is to strengthen and widen the European citizen science community by designing and implementing a variety of activities that serve to build capacity, raise awareness and establish sustainable regional, national and EU-wide citizen science networks. ECS aims to achieve impact in several dimensions, including in social/societal and political, scientific and institutional, as well as technological and economic impact areas.

In this context, social/societal impact covers both smaller-scale changes on an individual level and larger-scale changes through which citizen science might affect society. This includes for instance the creation and strengthening of communities, individual and collective empowerment, effecting changes in behaviour and attitudes, and supporting the acquisition of new knowledge and skills. Political impact refers to changes brought about on a policy and decision-making level. This entails the use of datasets, knowledge or other outputs developed by citizen science initiatives as an evidence-base to develop policies and to make policy decisions. On a science policy level structural changes in the academic systems and stronger support for open science and citizen science practices are considered under this impact area.,

Scientific impact focuses on the academic sphere and entails adaptations that allow for citizen science to thrive, integrating it into the processes of scientific innovation and knowledge production to create new and high-quality results in better alignment with societal needs. It is about engaging scientists and research organisations to achieve a greater acceptance of, but also a greater capacity to do citizen science and open science. Through this, human capital in R&I is strengthened and the diffusion of knowledge and open science is fostered. Institutional impact is about sustainable changes within organisations to enable and promote citizen science and public engagement in science and technology. It entails setting up clear institutional procedures, structures and regulations that allow for the consistent implementation of citizen science processes at a high level of quality. It also implies building up the necessary institutional expertise to do so.

Technological impact focuses on understanding technological improvements to respond to data gathering, analysis and dissemination of citizen science initiatives. It entails the involvement of citizens in technology development and enabling them to make

meaningful contributions, ensuring that technology is better aligned with societal needs and is accessible to a wider range of users. The impact of citizen science on technology is multifaceted, could empower individuals, promote collaboration, improve digital skills and literacy of participants, and foster technological advancements for the benefit of society as a whole. Economic impact, then, is centred around understanding the distribution of direct and indirect economic benefits that the implementation of citizen science can bring to organisations, communities, science, and society as a whole. This encompasses the examination of how citizen science initiatives contribute to economic growth, job creation, but also environmental sustainability, science education, and citizen empowerment from an economic perspective. The analysis of the economic aspects of citizen science brings insights into the tangible and intangible value it generates, both in terms of financial outcomes and broader societal benefits.

To achieve these impacts, ECS engages relevant stakeholders on all levels, from citizens to scientists (early career and established), to institutional decision makers and policymakers from the regional to the EU level. Our presentation showcases how we approach and structure the vast endeavour of evaluation and impact assessment of a comprehensive and multi-dimensional project such as ECS. We introduce the Logic Model Approach employed by the ECS impact assessment team to organise and visualise the concrete impact pathways pursued by the project and elaborate our ECS Metrics Matrix which operationalises the effects we seek to achieve – on R&I cultures, infrastructures, and the multiple stakeholders directly involved in ECS and other citizen science initiatives. Citizen science necessitates a broadening of our understanding of relevant impacts that can and should be achieved by R&I projects. New evaluation frameworks such as CoARA and SCOPE make important strides in rethinking evaluation beyond common academic metrics, opening it up to better grasp the impact of complex multi-stakeholder collaborations represented by the field of citizen science. At the same time, CoARA struggles to operationalize its call for qualitative evaluation in a way that covers the specificities of this field. SCOPE, while comprehensive and truly innovative, is not easily translated into the impact pathways logic currently favoured by the European Commission. At this intersection of existing evaluation structures and (comparatively) novel areas of application, our research elaborates various relevant measures of success, including an increase in knowledge, interest, and awareness, as well as more positive attitudes towards citizen science by all stakeholders of the quintuple helix. Adaptations of CS infrastructures and implementation of developed recommendations are other important outcomes of the project, all of which contribute to a set of wider impacts along the impact areas described above.

From there, we present some first insights, based on impact stories, self-assessment surveys and experience sampling, among other data gathering instruments. Special focus will be given to qualitative methodologies which are particularly well-equipped to provide evidence for the alternative impact metrics sought by citizen science projects, touching also on how we employed these instruments in a context-sensitive manner to scrutinise the effects of different project activities, formats, and stakeholder communities while also looking beyond the horizon of the ECS project to explore how impacts arise from the combined efforts of different stakeholders towards establishing citizen science in Europe. Finally, our contribution will show first results from our evaluation activities, focussing particularly on our achievements in terms of social and political impact, and argue that an integrated approach to evaluation and impact assessment is highly beneficial for project consortia to effectively work towards the goals to be achieved, especially in sprawling projects with vast interdependencies.

Lena Pfeifer
Leibniz-Centre for Agricultural Landscape Research (ZALF) e.V., Germany

UNFOLDING SOCIETAL IMPACT: APPLICATION OF THE SYSTEMIC RESEARCH IMPACT ASSESSMENT FRAMEWORK FOR THE DEVELOPMENT OF A FORMATIVE EX ANTE APPROACH

ABSTRACT

Recently, a new type of Research Impact Assessment (RIA) has emerged. The increased emphasis on societal responsibilities and the introduction of the Sustainable Development Goals (SDGs) have spurred the development and use of RIA approaches that assess research contributions to societal change and transformation. Over the past decade, an increasing amount of novel RIA approaches incorporate societal target systems, necessitating the inclusion of stakeholder perspectives and priorities to address trade-offs between societal goals. In response to this evolving landscape, we have developed a comprehensive RIA framework to support research in assessing and strategically planning its contributions to societal transformation.

Through the review and analysis of 70 RIA approaches the Systemic Research Impact Assessment Framework (Pfeifer & Helming 2024) summarizes four components aimed at enhancing societal impacts: (1) an integrated component facilitating reflection on impacts across all dimensions of sustainability, (2) a missions-oriented component aligning with societal goals to enhance societal relevance, (3) an inclusive component enabling broad participation to bolster the legitimacy of research and its impact, and (4) a strategic component for selecting suitable assessment scales and time dimensions to ensure effectiveness.

When applied in an ex ante approach, such systemic RIA can significantly aid strategic research planning. Within the LeNa Shape project, funded by the German Federal Ministry of Education and Research (BMBF, grant numbers 01UV2110F-G), we developed and applied such a formative ex ante RIA approach, executed in workshop series. We applied the Systemic Research Impact Assessment Framework in the design of the approach:

1. Integrated: In our approach we decided to assess impact on three impact levels, including all dimensions of sustainability as well as impact generating processes.

- a. Context-specific impacts (e.g. improved ecosystem functions)
- b. Societal Effects and Trade-Offs (environmental, social, economic and governance)
- c. Transformational Impact (contribution to the SDGs)

Although the workshops were developed in the context of natural resource use and management, the impact levels can be adapted for various contexts and research fields.

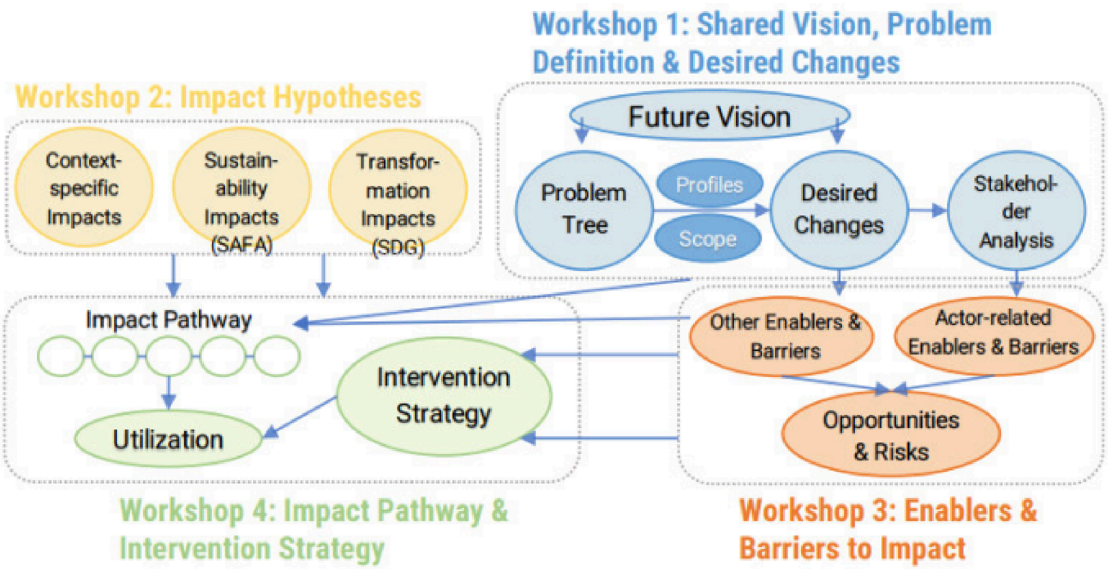
2. Mission-Oriented: The third impact dimension has a clear mission-orientation by its link to the Sustainable Development Goals (SDGs). However also the subsequent impact

dimension can be liked to missions, when appropriate. For example, for soil research we made use of the five soil functions formulated in the EU Soil Mission for our 1st level impact (context-specific impact). 3. Inclusive: In our approach we emphasize the necessity to include stakeholder perspectives in the selection of impact criteria and research topics. Implicit stakeholder oriented exercises (stakeholder assessment; actor related barriers and enablers) should be complemented with active participation of stakeholder groups. 4. Strategic: The aim of the developed approach is the formative analysis and planning of impact of future research. It was iteratively and adaptively applied, tested, and further refined in case studies. Through workshops, ex ante impact assessments were conducted at multiple scales: I) project level before grant application, II) project level after funding approval, III) program area of research institute.

The approach was iteratively applied and adapted in four case studies in agricultural research and tropical marine research. The resulting workshop series (Ferse & Pfeifer 2024) consists of 4 consecutive workshops, preparatory tasks and review (Figure 1):

- Workshop 1: The first workshop is dedicated to collaboratively defining the problem being addressed, the scope, and the common terminology of the research project or program. Additionally, the desired changes are outlined. The workshop concludes with a stakeholder analysis to prepare for subsequent workshops and to identify stakeholders for verifying the mapped impacts.
- Workshop 2: The second workshop examines the various potential impacts of the project or program, identifying the most relevant impact criteria and categories, and discussing possible indicators. By the end of the workshop (or shortly thereafter), an initial draft of the impact pathway is developed.
- Workshop 3: In the third workshop, participants will backtrack the necessary research actions, as well as identify relevant barriers and enablers needed to achieve the desired impacts defined in the first workshop. This session will help participants pinpoint key actors and institutions, reflect on enabling processes, recognize opportunities and resources, and assess risks associated with achieving the envisioned impacts.
- Workshop 4: In the fourth and final workshop, participants will finalize the impact pathway for the proposed research project or program. Additionally, they will outline a detailed intervention strategy and conceptualize how they intend to use the developed materials, whether for grant proposals, reflection, monitoring and evaluation strategies, communication, or reporting.

FIGURE 1 WORKSHOP OVERVIEW (FERSE & PFEIFER 2024)



The practical application of the developed ex ante RIA approach in four case studies (Table 1) has demonstrated its potential in anticipating the prospective impacts of research activities, along with its limitations and potential solutions. While the analysis and synthesis of the case studies are currently ongoing, preliminary results are as follows:

- Valuable tool for impact planning: Overall, the workshop series was perceived as a valuable tool for collaboratively planning impactful research. In particular, the creation of a joint understanding of the project's targets and anticipated pathways for achievement was considered helpful. Detailed feedback is currently being surveyed across the case study participants.
- Resource intensiveness: The workshop series required a significant time commitment from all participants. While most workshops were conducted online, in-person workshops would be preferable, albeit requiring additional financial and time resources for travel and workshop hosting. These required resources should be acknowledged.
- Moderation and facilitation: Effective facilitation and moderation should involve at least two persons. While participants do not need to be familiar with the concepts of societal impacts and impact planning, organizers and facilitators should have a solid understanding of the underlying concepts.

- Indicators: The workshops applied a criteria/indicator system to assess the impact of research, utilizing established indicator systems. Context-specific indicators, derived from ecosystem services assessments (CICES 2018; WRI 2015), help anticipate how research in natural resource management contributes to societal goals. For agri-/aquaculture and fisheries, social, environmental, and economic side effects are explained through SAFA indicators (FAO 2013), aiding in predicting sustainability impacts. Other contexts require tailored indicator systems like the Montreal Process Criteria and Indicators or GRI Standards. Lastly, SDG contributions are forecasted using criteria and indicators from the Global indicator framework (UN 2021). Using established indicator systems facilitated quick understanding of the criteria and ensured comparability.
- Stakeholder participation: While explicit stakeholder participation was considered essential, time restrictions and limited stakeholder availability only allowed for implicit consideration of stakeholder perspectives through stakeholder analysis. This is considered a limitation for the results of our case studies
- Utilization of workshop results: The workshop material was utilized in the grant proposal of the case study DAKIS 2. The grant proposal included the final impact pathway visualization and impact assumptions developed in the workshop series, leading to the approval of project funding. The case study SusWEF utilized the workshop material in their kick-off conference to gain a joint understanding of the project aims and planned activities.

TABLE 1 OVERVIEW CASE STUDIES

	Case Study 1	Case Study 2	Case Study 3	Case Study 4
Name	SusWEF	PaMo	PA Nutzung und Schutz aquatischer Ressourcen	DAKIS 2
Type	Funded Research Project	Funded Research Project	Program Area	Research Project before grant application
Participants				
Utilization	Use of material in Kick-Off conference	TBD	TBD	Integrated impact pathway in grant application > funding approved

The Systemic Research Impact Assessment Framework has proven to be invaluable for strategically designing a RIA approach. The experiences gained from applying this RIA approach in the case studies underscore the critical importance of prioritizing ex ante/formative Research Impact Assessment (RIA) with increased attention and support from funders and institutions. Given the resource-intensive nature of impact assessment, it must be taken into account in project funding and evaluation processes. Effective implementation requires support and incentives through favorable framework conditions, such as empowering individual researchers to conduct impact assessments and enhancing impact understanding and competencies within research institutions through supportive structures for impact reporting and support. Moreover, conditions for research funding and calls for proposals should be adjusted to incorporate impact-oriented requirements and provide support for RIA through seed/research support grants.

REFERENCES:

- Food and Agriculture Organization of the United Nations (FAO) (2013). SAFA Indicators: Sustainability Assessment of Food and Agriculture [Online]. Available at: https://www.fao.org/fileadmin/templates/nr/sustainability_pathways/docs/SAFA_Indica t ors_final_19122013.pdf [Accessed 30.01.2024].
- Ferse S.C.A., Pfeifer L. (2024) Ex ante Impact Pathways. A guide to conducting workshops for impact planning. Prepared as part of the LeNa Shape project. Leibniz Centre for Tropical Marine Research (ZMT) Bremen GmbH and Leibniz Centre for Agricultural Landscape Research (ZALF) e.V. M¸ncheberg, Germany. https://doi.org/10.21244/zmt.2024.001
- Haines-Young R (2023) Common International Classification of Ecosystem Services (CICES) V5.2 and Guidance on the Application of the Revised Structure. Fabis Consulting Ltd., Nottingham, UK.
- Pfeifer, L., and Helming, K. (2024). Effective mission-oriented research: A new framework for systemic research impact assessment. Research Evaluation 00, 1-125. https://doi.org/10.1093/reseval/rvae003
- United Nations (UN) (2021). Global Indicator Framework for the Sustainable Development Goals and targets of the 2030 Agenda for Sustainable Development [Online]. Available at: https://unstats.un.org/sdgs/indicators/indicators-list/ [Accessed 03.01.2022]
- World Resources Institute (WRI) (2014). Indicators of Sustainable Agriculture: A Scoping Analysis [Online]. Available at: https://www.wri.org/data/indicators-sustainable-agriculture-scoping-analysis [Accessed 30.01.2024]

Examining Instruments in European Research and Innovation Policy

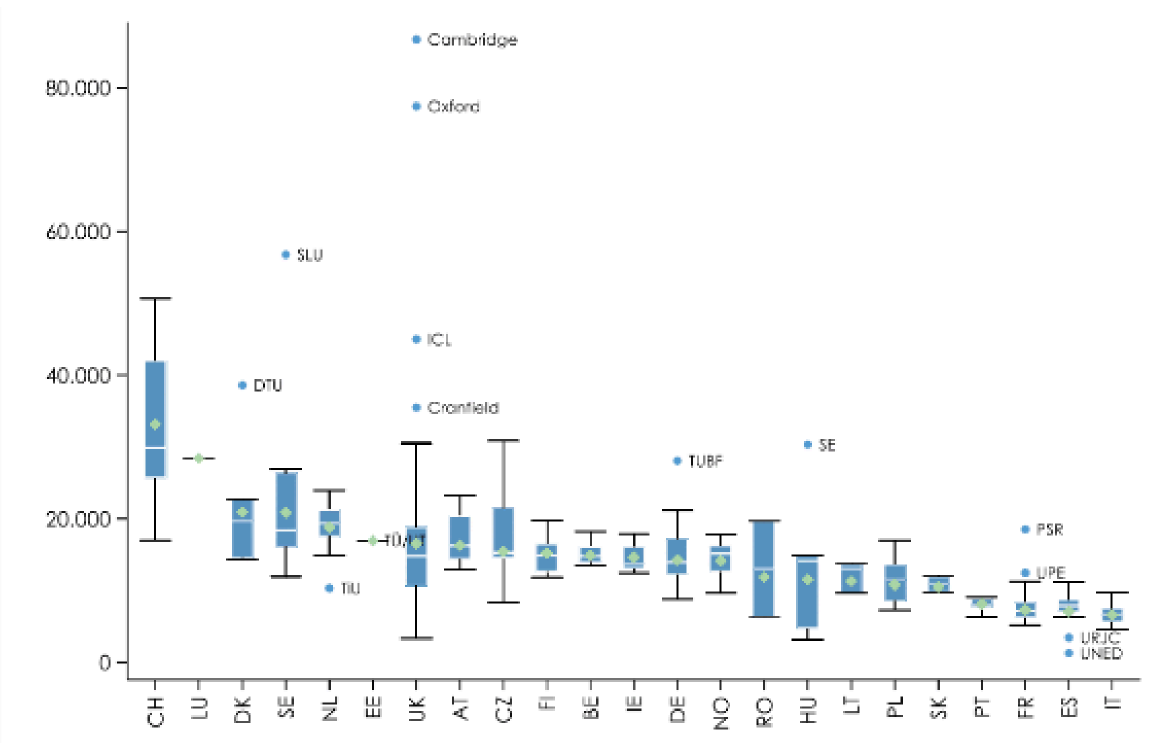
Kathrin Hofmann
WIFO, Austria

UNIVERSITY RESEARCH PERFORMANCE: ALL 'BOUT THE MONEY?

ABSTRACT

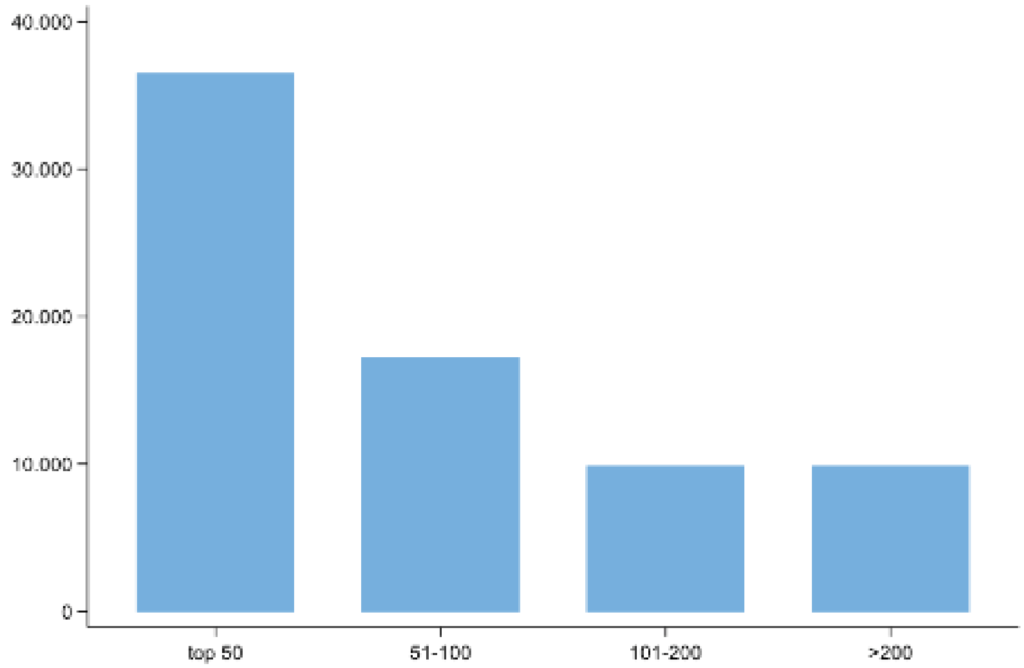
This paper starts with surveying the main strands of the literature relevant for institutional knowledge production and the puzzling observation that while some empirical contributions do not pay any attention to the role of the level of monetary resources or expenditures, others focus almost uniquely on money as an explanation for differences in university research performance. Casestudies and more qualitative accounts, however, clearly point to both differences in organisation and governance, giving rise to different incentives and managerial leeway, and differences in monetary resources as characteristics of favourable conditions for knowledge production. We build upon the publicly available ETER dataset and match it with CWTS Leidenranking data on university knowledge production to create the—to the best of our knowledge—largest database repository for investigating determinants of university knowledge production. Descriptive accounts of the data indicate that there are large differences in both knowledge production and monetary resources within and across countries. Top knowledge-producing institutions seem to spend disproportionately more than institutions outside the top 100 or so. Our dataset would also allow for studying more indetail vertical stratification of university systems, but we leave this for a different paper.

FIGURE 1: TOTAL CURRENT EXPENDITURES PER STUDENTACROSS UNIVERSITIES AND COUNTRIES, 2020



Source: ETERproject. Download date 30.3.2023, augmented by own efforts.

FIGURE 2: EXPENDITURES PER STUDENT PERLEIDENRANKING RANKGROUP, 2020/2018-2021



Source: ETERproject. Download date: 30.3.2023andCWTSLeidenranking.

To estimate the determinants of institutional knowledge production, we propose to use linear dynamic panel data models with a lagged dependent variable, as current knowledge production is influenced by existing knowledge stocks and well-documented cumulative advantage in science gives rise to persistent knowledge production patterns. To control for size, we argue in favour of using the number of students.

Our results clearly indicate that while expenditures controlling for size are very important for knowledge production, more so at the upper end of universities, they are not the only relevant factor—the coefficient on expenditures per student is clearly below 1. Just pouring money into an institution without appropriately designed organisational and governance structures is unlikely to produce the desired results. At the same time, the enormous differences in expenditures controlling for size outlined in Section 3—by a factor of 3 even between universities in countries of similar income levels, such as the UK vs. Germany—point to the need for significantly increasing monetary resources if European universities want to catch up to leading knowledge production levels. High levels of knowledge production in universities need both types of production factors.

TABLE 1: REGRESSION RESULTS–BASELINE SPECIFICATION.

Top10-Publications per student	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
Lagged dependent variable	.699	.085	8.24	0	.533	.866	***
Number of students	-.427	.095	-4.50	0	-.614	-.241	***
Number of staff	.11	.054	2.05	.04	.005	.216	**
Expenditures per student	.199	.024	8.19	0	.151	.247	***
Share of third party funding	.001	.004	0.36	.716	-.006	.009	
Share of students in FOE1-4	.006	.002	2.39	.017	.001	.011	**

Mean dependent var	-4.460	SD dependent var	0.952
Number of obs	2302	Chi-square	78935.643

**** p<.01, ** p<.05, * p<.1. All numbers in logarithms except for shares. FOE 1-4 – fields of education which are not equipment intensive. FOE1-4 – non-expensive Fields of Education such as social sciences.*

Our paper is a work in progress, and our results are limited so far. The question remains how much non-monetary factors matter compared to monetary resources. In future extensions, we plan to address this. Our current results are however already highly relevant for national and European-level policies to improve the conditions for knowledge production in universities, often termed excellence initiatives." Such policies should draw on the available evidence on the determinants of knowledge production. Our survey of the literature has shown that different contributions emphasise different factors. In a highly stylised summary, some point to differences in incentives and organisational structures conditioned by knowledge production. Implementing organisational changes does not necessarily need a lot of money. Others take the opposite view, singling out money as the key factor for making it to the top in rankings. Our results don't make it easier for policies, as essentially, we are saying that both are necessary.

Current initiatives to boost European universities usually fall far short with respect to the gaps in budgets observed. The European Universities initiative, e.g.—to be fair, not yet an excellence initiative strictly speaking—provides 2 million € per selected university alliance. The German Excellence Initiative provides additional funding to universities selected as "excellent" in the range of up to 15 million €. Given the differences in expenditure levels outlined, the German funding would have to be 20 times higher per university. This could only work if the existing resources of the excellence initiative were not spread thinly over about 30 universities but concentrated on two to three universities. It would be very difficult to fund a large share of the university sector at the level of top productive universities. The same holds true for the European Universities initiative or any new European Excellence initiative. Currently, 60 alliances involving 500 higher education institutions are planned by mid-2024.4 Spreading funds thinly over such a large number of institutions is unlikely to produce measurable effects. At the same time, the existing initiatives rarely ask for structural reforms to organisation and governance of universities—universities are mostly selected for the research quality in so-called clusters, e.g., as in the German excellence initiative. Effective policies would combine both, concentrating large amounts of additional funding on a few universities conditional structural reforms that optimise the conditions for knowledge production. If implemented, such policies would more broadly boost the EU's capability to catch up to the frontier in knowledge production, at a time of increased geopolitical tension and efforts to exploit unilateral technological dependencies.

REFERENCE:
https://education.ec.europa.eu/education-levels/higher-education/european-universities-initiative

Hendrik Berghäuser
Fraunhofer Institute for Systems and Innovation Research, Germany

EVALUATING A MOVING TARGET: INSIGHTS FROM THE ONGOING EVALUATION OF THE FEDERAL AGENCYFOR DISRUPTIVE INNOVATION IN GERMANY (SPRIND GMBH)

ABSTRACT

The promotion of disruptive innovation has received increasing attention in innovation policy in recent years. One reason for this is accelerated innovation processes due to digital transformation of entire areas of society and the emergence of new competitors, for example in Asia. Economies are therefore faced with the challenge of developing ideas for disruptive innovation in order to ensure long-term competitiveness and economic growth. Another important reason for the growing relevance of radical innovation in innovation policy is an increased political focus on solving societal challenges. Due to their high innovation degree and considerable market effects, disruptive innovations—unlike incremental innovations—have a supra-regional impact and can thus make a significant contribution and help to shape social transformation processes, for example through new mobility concepts, new treatment methods for diseases, or new solutions on the path to climate neutrality. As a consequence, various initiatives in Europe can be observed that target the development of new funding approaches. Also, new funding agencies to promote disruptive innovation have emerged. These funding bodies follow an ambitious organisational model and develop funding and financing instruments that are specifically designed to actively support disruptive technologies and radical innovations with a significant social and/or economic impact at an early stage. The funding instruments are characterised by a high degree of openness towards technological and methodological approaches as well as a strong focus on problems and applications. They also aim for a high degree of scalability and require a new willingness to take risks. At the same time, the tension between ensuring the greatest possible organisational independence on the one hand and the use of public funds for high-risk projects on the other makes the creation of such a funding agency an extremely complex undertaking in political, legal, and organisational terms.

In Germany, the Federal Agency for Radical Innovation (German: Bundesagentur zur Förderung vonSprunginnovationen SPRIND) was founded in December 2019 with the Federal Republic as the sole shareholder. In addition, the current coalition between Social Democrats (SPD), Liberals (FDP), and Greens planned a new law—the so-called SPRIND Freedom Act-in order to improve legal and financial conditions for SPRIND's mission to promote radical innovation. This law was elaborated in the second half of 2023 and came into force on December 30th, 2023.

Currently, SPRIND is offering a mixture of top-down and bottom-up funding approaches. The three main instruments are:

- validation contracts (to validate the potential for disruptive innovation);
- spinning our subsidiary firms (R&D companies);
- challenges (participants try to develop solutions for a given problem).

The SPRIND Freedom Act, which has now come into force, gives SPRIND the opportunity to offer additional financing instruments, including private-sector financing instruments such as investments in existing companies. These instruments are currently still under development.

In spring 2023, the German Federal Ministry for Education and Research (BMBF) published a call for tenders for the evaluation of the SPRIND. The aim of this evaluation is to analyse and assess SPRIND's funding approach, its process efficiency, its governance, and the effectiveness of its funding instruments. The evaluation shall comprehensively assess the structure of SPRIND since its foundation (approaches practiced up to date) and its operation in the further course, while taking into consideration further developments, in particular the planned legal changes with regard to the SPRIND Freedom Act. The evaluation is divided into five substantive work packages.

1. Evaluation of the overall SPRIND concept (assessment of the organisational concept and the current funding instruments)
2. Evaluation of the planned changes (in particular with regard to the SPRIND Freedom Act)
3. Evaluation of organisation and equipment, including process efficiency
4. Evaluation of the staffing concept and management
5. Evaluation of cooperation and environment (including embedding SPRIND in the innovation system)

The evaluation runs from July 2023 to December 2024.

The evaluation description results in complex requirements and challenges for the design of the evaluation:

- The uniqueness of the Federal Agency as an object of evaluation makes a counterfactual analysis impossible and complicates reliable comparisons with other examples.
- The analysis of newly established organisational processes and funding instruments underconditions that may still change in the course of the evaluation with regard to the SPRINDFreedom Act places particular demands on the evaluation approach (evaluating a "moving target);

- The interaction of different funding approaches with different program logics and the corresponding long funding periods for the generating radical innovation requires an evaluation approach that goes beyond simple input-output-outcome-impact chains.
- As the majority of the interventions have not yet been completed and the selected projects are still being funded, it is difficult to measure longer-term effects.
- The comprehensive evaluation of SPRIND as a new player in the German innovation system also requires it to be located in the broader innovation policy funding portfolio of the federal government and the international trend of developing funding approaches for radical innovations.
- The tendered evaluation is a combination of program and organisational evaluation, which requires a particularly broad methodological approach.

Due to the great importance of external conditions and influencing factors on effects, we rely on atheory-based evaluation. This approach is based on well-founded statements about the impact paths and mechanisms of the various funding instruments and interventions. The evaluation approach examines possible effects primarily in terms of which factors have contributed to an observed change, without assuming explicit cause-and-effect relationships. With regard to the explicit goal of SPRIND to induce breakthrough innovations, we propose to trace so-called "impact pathways" as part of the goal achievement and impact measurement.

Methodologically, the evaluation approach relies on a triangulation, which is based on various quantitative and qualitative methods of data collection and data analysis. In particular, this involves reactive methods in order to link the assessments of different stakeholders with each other in the sense of a multi-perspective approach. These are supplemented by non-reactive methods, which are intended to tap into the extensive range of documents and key data regarding the governance as well as the activities of SPRIND. In addition, the evaluation will identify international benchmarks for the generation of radical innovation in order to arrive at valid evaluation results. Specifically, the evaluation approach relies on the following methods:

Evaluation method	Purpose
Documents analysis	Assessing governance-relevant documents to get a better understanding of the evaluation object and its funding activities
Literature analysis	Comprehensive analysis of relevant literature on radical innovation for an empirically based evaluation of existing funding approaches and indicators.
Standardized interviews	Interviews with various policy actors (ministry officers responsible for SPRIND governance, members of parliament), SPRIND representatives (i.e. SPRIND management, but also members of the Advisory Board), interviews with innovators funded within the three streams of funding approaches, interviews with external national and international experts) in order to get a holistic overview and insights for the evaluation
Online Surveys	Survey among the participants of the challenge as well as an online survey among SPRIND employees for assessing the efficacy and effectiveness of the funding instruments as well as the working conditions at SPRIND.
Process Analysis	This analysis forms the descriptive-analytical step, based on the document analysis and first interviews with SPRIND representatives, in order to map all bodies, functional units, their responsibilities.
Media Analysis	Carring out a media analysis to evaluate the external impact of SPRIND's communication activities
Cost-effectiveness analysis	Analysis to assess the cost-effectiveness of implementation and measures.

In addition, a law firm is also involved in the evaluation in order to carry out a legal examination, as many evaluation questions have legal implications.

As the Federal Ministry of Economics (BMWK) and the Federal Ministry of Finance (BMF) are also involved in the management of SPRIND in addition to the BMBF, and a federal law to improve the financial and legal conditions for the promotion of disruptive innovation is planned during the evaluation period, the evaluation needs to ensure a high level of sensitivity with regard to political communication and the involvement of various political actors from legislative and executive. Various political actors are attempting to use findings from the evaluation for political purposes. Numerous enquiries and invitations as experts in legislative consultation processes on the SPRIND Freedom Act lead to tensions between maintaining neutrality on the one hand and providing political advice on the other. In addition, the evaluation of a funding institution (SPRIND) on behalf of another institution (BMBF) requires open, trust-building, and transparent communication.

Susanne Buehrer, Sarah Seus, Florian Wittmann
Fraunhofer Institute for Systems and Innovation Research, Germany

A NEW APPROACH TO OPEN THE BLACK BOX FOR THE IMPACT ANALYSIS OF A MIS-SION-ORIENTED RESEARCH STRATEGY

ABSTRACT

Our article presents a new conceptual approach for the impact analysis of an ambitious mission-ori-ented research funding strategy (FONA). The FONA strategy is divided into three main strategic objectives (achieving climate goals; researching, protecting, and using habitats and natural resources; developing society and the economy—good living conditions throughout the country); eight priority fields of activity (for example, avoiding and reducing greenhouse gases (mitigations), securing natural resources (water, soil), and innovative regions); and 25 specific actions (https://www.fona.de/en/fona-strategy/in-dex_en.php).

The FONA strategy is a heterogeneous and multi-measure policy instrument aiming to link research activities with broader societal (sustainability) goals. Under the FONA strategy, different types of research projects (i.e., traditional collaborative research projects, interdisciplinary and transdisciplinary research) are funded, but also research infrastructures, support for young scientists, and different studies providing policy intelligence. FONA funds projects in a wide range of topics under the umbrella of sustainability research. Sustainability research is not a clearly defined scientific discipline; its main two aspects that are connecting different topics are, however: first, scientific research that is aligned to societal problems and/or contributes to societal transformation processes; and second, how research is conducted with a focus of inter- and transdisciplinary processes (Komi-yama and Takeuchi 2006; Gallopin et al. 2001). Our first step was to get hold of this heterogeneity by elaborating the specificities of FONA-funded research and using these specificities as outcome dimensions in an overall impact pathway of the FONA strategy. We defined six "FONA-dimensions."

A) contribution to knowledge development in the field of sustainability research;

B) international cooperation;

C) interdisciplinary and transdisciplinary research;

D) integration of a systemic perspective into the research process; technology transfer (commercial use);

E) creating practical knowledge for scientific application.

Taking into account this heterogeneity and multitude of objectives that FONA activities are aiming at, a traditional evaluation approach—developing impact pathways for FONA’s 25 main objectives (“Ak-tivitäten")—returned"not to be expedient. To get hold of the

complexity, we needed an intermediate step in the form of an overarching approach in order to achieve better systematisation and comparability between the FONA-funded activities. This intermediate step consists in the development of ideal-typical theories of change (ToCs) that describe the central impact logics of different policy instruments. Assuming that different policy instruments pursue similar impact mechanisms despite different thematic focuses, the different impact logics can be described in a comparable way that later can be tailored to thematic areas/fields of action. This approach offers two key advantages. Firstly, this approach allows a systematic aggregation of FONA activities at different levels (programs, actions, fields of activity, FONA strategy). At the same time, the policy instrument ToCs can be understood as modular building blocks of the individual fields of action, which bundle different policy instruments and can thus be described, compared, and further developed in a simple and systematic way. In the case of the FONA strategy, we are aggregating the effects of the policy instruments toward six main FONA-specific dimensions.

There are already established classifications for policy instruments in the literature (see, for example, Hufnagl 2010, OECD STIP Compass 2021). In the following, we orientate ourselves on the comprehensive typology of the OECD STIP Compass, which distinguishes between 28 policy instruments. These 28 policy instruments represent the starting point for a review of their relevance to the FONA strategy. On this basis, a total of 10 policy instruments were identified as central, with additions and differentiations being made where this seemed necessary, in particular as FONA is a research funding program, in comparison to the OECED-typology covering the whole innovation process. This applies in particular to the area of project funding in order to do justice to the collaborative projects funded by the BMBF, so that further distinction was made for collaborative research projects: we distinguish:

a) collaborative projects only with researchers from a research organisation;

b) collaboration researchers with industry partners;

c) transdisciplinary research projects with researchers and partners from (local) administrations and civil society.

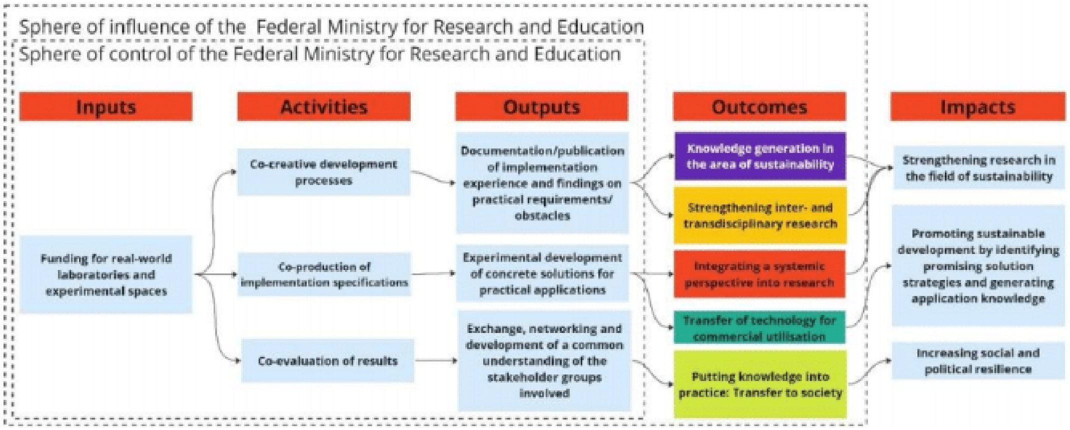
The central policy instruments identified for FONA were also supplemented by the increasingly important real-world laboratories/experimental spaces, as these are not yet included in the STIP Compass. The following policy instruments were thus identified as central FONA instruments:

1) Collaborative research within science; 2) Collaborative research (with industry, private sector);

3) Collaborative research (with municipal or social actors);
4) Junior researchgroups;
5) Network structures & collaborative platforms;
6) Research infrastructures;
7) Science and innovation competitions;
8) Strategies, agendas, plans,
9) Policy intelligence (e.g., evaluations, accompanying research, foresight activities);
10) real-world laboratories and experimental spaces

For each of the selected policy instruments with FONA relevance, theory-based theories of change were developed on the basis of scientific literature, which describe the central interdependencies along the inputs, activities, outputs, outcomes, and impacts and also take into account the various possible effects of an instrument. The novelty of our approach is that we identified specific mechanisms leading to effects (conditions for success and hindering factors) for each policy instrument. These are based, on the one hand, on the theoretically derived categories for conditions for success and, on the other hand, embedded in them, instrument-specific requirements that were identified on the basis of scientific literature. In the last step, we analysed all mechanisms of each instrument in order to come to one classification of “success mechanism." We have identified three types of conditions for success (support design, interaction and reflection processes, project-external factors/context). In the area of interaction and reflection processes, we focus in particular on concepts such as RRI (van Mierlo et al. 2010, Peter et al. 2018, Ravn et al. 2015, Stilgoe et al. 2013; Seus et al. 2023), the LeNa reflection framework (Daedlow et al. 2016, Ferretti et al. 2016), and considerations on productive interactions (Spaapen und van Drooge 2011, Spaapen et al. 2011).
The impact pathway is presented for each policy instrument in the form of an overview (see an example in figure 1) and an accompanying text of approximately three pages, including the description of policy instruments individual “success mechanisms." The well-elaborated descriptions lay down the theoretical basis of the impact pathway and allow not only to show the intended effect chains but especially to lay down the relationship between effect categories and conditions for effect development. Based on a literature review, we thrive to establish for each policy instrument a state-of-the-art theory of change.

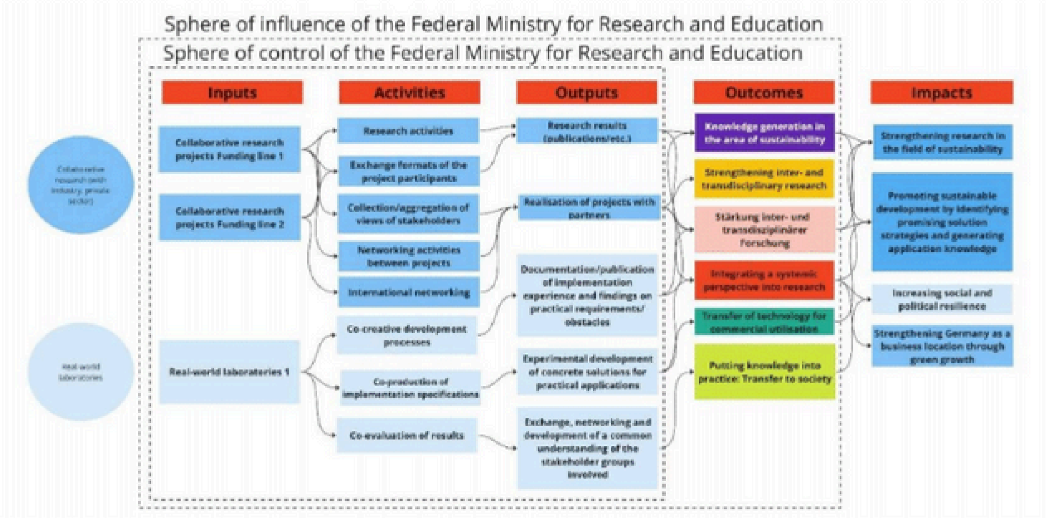
FIGURE 1: ILLUSTRATION OF THE I-A-O-O SCHEME FOR REAL-WORLD LABORATORIES



Source: Own illustration

In view of the modular nature of the individual policy instrument ToCs, particular attention was paid to comparability between the various policy instruments. The key FONA dimensions, which can be found in various combinations in the individual policy instruments, are the central fixed point in the presentation. Building on the identified policy instruments (ToCs), it is possible to develop specific but still comparable theories of change for individual FONA actions. The result is a specific ToC for a selected FONA action, as the following example shows:

FIGURE 2: EXAMPLE OF COMBINING POLICY INSTRUMENTS TOCS INTO AN ACTION TOCS



Our approach is currently developed and applied to several FONA fields of activity. We will discuss in our presentations our experience with this new approach and discuss the usability and effectiveness to assess the effectiveness of a research strategy using 10 policy instruments as a proxy.

Dr. Julian Emmler, Sophie von Knebel, Laura Barking, Dr. Felix Gaisbauer, Frida Salge-Peters
DLR Projektträger, Germany

NAVIGATING EVALUATION COMPLEXITIES: INSIGHTS FROM ASSESSING EUROPEAN FUNDING PROGRAMMES

ABSTRACT

The case studies presented in this abstract contribute to Strand 3 (Examining instruments in European research and innovation policy) and, in particular, to the assessment of EU funding programmes and the identification of lessons learned for future programmes. To this end, we use two examples for projects that demonstrate the challenges in monitoring and evaluating national contributions to EU funding programmes. By examining these challenges in greater detail, we can identify lessons learned from those evaluating and monitoring efforts, which will in turn improve the assessment of EU funding programmes beyond the presented case studies.

European funding programmes play a crucial role in promoting economic development, social cohesion and environmental protection in the European Union. Therefore, it is important to continuously monitor and evaluate these programmes to ensure that the funds provided are used effectively and that the desired objectives are achieved. Moreover, these evaluations can help to identify potential for improvement and with this contribute to even better targeted programmes. One challenge, however, is the necessary interaction between the EU and the national levels for the further development of the programmes. The aim of the present investigation is therefore, to reflect on the evaluation processes of EU funding programmes and, additionally, to reflect on methods that are used in these processes and how these have already or will have to adapt to our fast-evolving European funding world. We do this by looking closely at two evaluation projects which we have conducted throughout the last year: The "Mid-term evaluation of the Erasmus+ programme generation 2021-2027" and the monitoring of the German projects funded through the "Important Project of Common European Interest - Next Generation Cloud Infrastructure and Services" (IPCEI- CIS).

Mid-term evaluation of the Erasmus+ programme generation 2021-2027

The first project we will use as a case study is the "Mid-term evaluation of the Erasmus+ programme generation 2021-2027", which we conducted in 2023/2024. The Erasmus+ programme aims to promote lifelong learning, enable sustainable growth, strengthen social cohesion and European identity and drive innovation. The aim of the project was the interim evaluation of the four education sectors (vocational education and training, adult education, higher education, school education) of the current Erasmus+ programme period 2021-2027 at the national level. The evaluation was intended to inform

the design of the next programme period, which will commence in 2027. This evaluation is characterised by the fact that the EU Commission provided a fixed catalogue of questions to be answered, which remains consistent across all participating countries. The catalogue consisted of 36 questions centred around the following criterions: effectiveness, efficiency, relevance, coherence and European added value.

The evaluation of the Erasmus+ programme presented several challenges, some of which were clearly related to the European nature of the programme. First, the timeframe was ambitious and rigid. There was no room for negotiation as the deadline was set at the European level, while the evaluation was commissioned by the responsible national authority, namely the Federal Ministry of Education and Research (BMBF) and the Kultusministerkonferenz (KMK). Second, the fixed set of questions mentioned above, which was designed to ensure the comparability of evaluations across countries, did not allow for questions specific to the German context. Third, access to secondary data collected through EU instruments proved to be problematic due to data protection concerns, requiring the involvement of several parties. This lengthy process prevented us from using the data within the set timeframe. Fourth, in the absence of access to secondary data, we had to rely more heavily on the accounts of the implementing organisations, which may have a vested interest in presenting their organisations in a positive light. Fifth, there are numerous target groups for the Erasmus+ programme, some of which are difficult to reach, both for the programme itself and for the evaluation team. Sixth, targets are set at the EU level. Particularly when targets are set in absolute terms, it is difficult to break them down to the national level. Finally, in the absence of data and with limited time, it is difficult to ascertain whether observed effects can be attributed to Erasmus+.

In light of these challenges we would like to discuss the following questions:

1) How to bridge the gap between the need to take account of the specific characteristics of the countries in the design of the questions and the need for comparability of the evaluations across countries?

2) How to deal with the need to rely on experts in the field to provide information, without simply and without critical analysis becoming an advocate for the interests of the implementing agencies? To what extent can the descriptions of the actors involved be applied at the national level without further investigation?

3)How to motivate non-participants or participants who have participated a long time ago (if they can even be reached) to take part in the evaluation? And how can an increased outreach effort be achieved within the limited time available?

4) How can EU targets be translated to the national level?
5) How can the intended impacts be linked to the programme, given the often-limited data available and the time constraints?

IPCEI-CIS

The second case study concerns the monitoring and evaluation of the German projects funded through the IPCEI-CIS. The IPCEI combines institutions and firms from 12 EU member state countries and aims at establishing a common European Cloud-Edge Continuum. Due to the integrated nature of the programme, its success is strongly dependent on the co-operation and dependencies of the different individual projects funded through the IPCEI-CIS.

From an evaluation perspective, this poses multiple challenges when it comes to meaningfully monitor and setting up the evaluation of a collection of individual projects, i.e. the German contribution to the IPCEI-CIS, which are embedded in a larger funding context. First, the M&E system for the German projects had to be developed before an overarching European M&E system for the whole program was in place. Second, judging the success of projects or the funding in Germany depends on the success of the whole IPCEI, which cannot be evaluated within the scope of the monitoring and evaluation mandate. Third, due to the highly dependent nature of projects and the multitude of involved stakeholders, many contextual factors which are difficult to measure for the national monitoring, will influence the monitored projects' development. Fourth, the complex target system of the funding program and the large number of involved stakeholders may lead to substantial changes in the structure or focus of (parts of) the program, to which the monitoring system has to react in a flexible and agile way.

In light of these challenges we would like to discuss the following questions and describe our approach to address them in the devised M&E system, as well as present lessons learned:

1). How to devise a country-specific M&E system that is adaptable to a not yet existent overarching M&E system, especially no overarching Theory of Change for the whole program?
2) How to monitor and later on evaluate the success of individual projects when there are many inter-project dependencies?
3) How to deal with challenges posed by timing issues (e.g. scattered project starts, implementation of program steps taking time due to complex program structure etc.), or by changing in the program structure (many stakeholders with different interests)?

4) How to report and integrate the different needs and requirements of stakeholders (EU commission, ministries, governing bodies of the program)?

Comparing the two evaluation systems

So far, we have focused on the individual challenges we encountered in the case studies. There is some overlap, and it is likely that these common challenges are representative of the difficulties that are encountered when monitoring and evaluating national contributions to EU programmes in general. Examples of such common challenges include the fact that the overall success of a programme cannot be determined by the success of the national contribution alone, but depends on the success of the programme as a whole. Another challenge in both case studies was the large number of stakeholders involved – hence, effective stakeholder engagement is critical to both projects.

However, it is important to note that there are significant differences between IPCEI-CIS and Erasmus+. A key difference is that IPCEI-CIS aims to establish an ongoing monitoring system and prepare an evaluation, while Erasmus+ is designed to produce a one-off evaluation. This helps us to highlight both, the challenges of devising an adaptable monitoring system without a pre-existing overarching framework, and the challenges to fit M&E systems in existing and potentially rigid overarching frameworks. In addition, the monitoring of IPCEI-CIS can mostly rely on standardised metrics and data sharing mechanism and the challenge lies in relating these to the progress of the program as a whole. In the case of Erasmus+, to measure the promotion European values and principles, an important challenge is to operationalise and measure attitudes in a meaningful way, as well as to include qualitative data for capturing the broader societal impacts.

As a result, the two case studies allow us to discuss a broad range of challenges attached to monitoring and evaluating national contributions to EU-programmes. By confronting these challenges head-on and devising adaptive methodologies, we contribute to the advancement of evaluation practices within the realm of EU funding programmes. Our efforts underscore the importance of continuous innovation in evaluation methodologies to ensure the effectiveness and relevance of future programme assessments.

Erik Arnold, Tomas Åström
Technopolis, United Kingdom and Sweden

CAN OLD DOGS LEARN TRANSITIONAL TRICKS?

ABSTRACT

While there is strong and increasing international consensus on the need for policies address sustainability and other societal challenges through socio-technical transitions, there is less clarity about how to do this based on existing state funding organisations and policy instruments. A common pattern is that authorities take traditional policy instruments and try to re-purpose them towards more transitional goals. Prominent examples include the German High-Tech Strategy Missions and the Dutch Top-Sector Missions, but the tendency is much wider spread (Larrue, 2021).

Our paper reports on experience from the seventeen Swedish Strategic Innovation Programmes (SIPs) launched in 2013-2017, identifying opportunities and obstacles for re-using what was originally a rather traditional funding instrument to more transitional objectives (Åstrˆm & Arnold, 2023). They provide a large-scale experiment from which we can draw conclusions about opportunities to build on existing organisations, governance and practices, and the extent to which bigger reforms are needed to support transitions. The SIP programme is, of course, only one example, though it is a rich one with seventeen different cases. So, our meta-evaluation both generates potentially useful lessons, but also a body of evidence that can be used in cross-country comparisons now that some of the other similar interventions internationally start to mature enough that lessons can be drawn also from them.

Sweden has had a long tradition of state-funded technology R&D programmes aimed at strengthening specific branches of industry and to an important extent steered by representatives of the respective branches and in many cases built on specific agreements between the branches and the government. In 2012, the government coalition's Research Bill called for these to be replaced by a new and more open funding instrument, open to all branches of industry and, almost as an afterthought, said that these should address not only industrial competitiveness but also the societal challenges, and be funded by three state agencies: Vinnova, Formas and the Swedish Energy Agency. The resulting SIP instrument funded a competition to establish public-private innovation consortia for up to 12 years. Each SIP was governed by an industry-dominated board, developed, and implemented an innovation agenda, established a programme office, and defined calls for project proposals to support the agenda. The agencies, issued the calls for proposals, assessed, and funded he proposals and funded the projects, avoiding the risk of adverse selection by the SIPs themselves. If all the SIPs last the course for the full 12 years, total funding will be some €1.45 billion, 45% from the public purse, and the balance in co-

funding from participating organisations.

There were 4 national SIP competitions, with 5 SIPs launched in 2013, 6 in 2014, 5 in 2015, and an additional one in in 2017. The first wave of 5 SIPs resembled repackaged branch research programmes, largely driven by industrial-academic groupings with long histories of organising branch programmes. Successively, however, the SIP waves became more socio-technical and more focused on sustainability, with the first wave tackling areas such as mining, metals, and manufacturing technologies, and the final one on Viable Cities. To some extent, their innovation agendas also evolved to become more concerned with sustainability through their lives.

Each SIP has been evaluated after 3, 6 and 9 years, as a condition for receiving continued funding. Our paper provides a meta-analysis of the 6-year evaluations, which are more detailed than the others. They provide a set of observations using consistent methods (which would not have been possible if we had mixed them with 3- and 9-year evaluations). The 6-year evaluations used mixed methods, on the one hand to deliver information about the SIPs' performance as technology programmes to support re-funding decisions, and on the other hand to observe their development through the lens of transition theory, as a basis for policy learning. Our focus here is mainly on the latter.

Based on a literature review in 2019, at the start of the 6-year evaluation, we developed an analytic framework, first using the 'functions' that the technological innovation systems (TIS) literature treats as necessary for developing new TIS (Bergek, et al., 2007) (Hekkert, et al., 2007). We added further functions from the transition management and niche management literatures, partly to extend the range of functions that need to be performed in technology and society and partly to incorporate 'management' functions needed in the policy instrument(s) used and in wider policy. This framework remained constant through the four years of the study, so that all the SIPs were viewed through the same lens. We observed the ways the extended list of functions were performed at the time of each six-year evaluation.

Based on this analysis, we classified the SIPs into three categories, with the emphasis shifting over time towards sustainability-related topics and composition.

- 10 reinforcers, , with a long history of branch programmes or similar instruments, and whose primary focus was on industrial competitiveness. These mostly started in the first two waves
- 5 transformers, more focused on sustainability transitions in areas such as transport infrastructure, smart buildings, and circular economy, from the third and fourth waves

- 1 TIS-builder, working to build supply chains and an ecosystem around the use of graphene
- 1 mixed case, which started as a reinforcer then radically changed course to become a transformer, aiming to change healthcare delivery systems using Internet-of-Things technologies

The transformers and the TIS-builder were more active in using the transformative functions in our framework than the reinforcers. The reinforcers tend to be rooted in mature industrial branches with oligopolistic market structures. The transformers are more based in functions or services, such as circular economy, urban planning. Big companies dominate the industrial contribution to all but 2 SIP (which are based in branches that are traditionally SME-dominated). Despite being incumbents in existing technological regimes, many of the big companies are increasingly interested in R&D to support transformative change, generally because they see that their markets will change and that it is therefore in their best interest to lead that change.

Our analysis of the way the SIPs used our long list of transitional functions showed that there were three clear categories of function

- Functions such as knowledge development and diffusion that are well established and traditional, and that can easily be done using the SIP instrument
- Functions such as creating arenas for priority setting, visioning, establishing directionality in R&D and reflexivity that could be done within the legal and organisational framework if traditional branch or technology programmes, but which could have been much more extensively done in the SIPs. This indicates that there is a great deal of scope for moving a fairly traditional innovation-orientated funding instrument towards tackling sociotechnical transitions
- Functions such as resources mobilisation, market formation, and creative destruction (ex-novation) that are outside the legal powers of the funding agents and therefore of the SIP instrument. These aspects clearly require changes in governance, the inclusion of non-R&D actors and other stakeholders if they are to be done

We have been in dialogue about our results with the funding agencies throughout this rather long project and have thereby contributed to the design of Impact Innovation (successor to the SIPs) and to their thinking about other programme designs.

Our immediate next step, ahead of the REvaluation conference, is to explore whether there is comparative information available about similar programmes from secondary sources, to provide a more international context, building on a background study we completed for the OECD about a year ago. We intend to work this conference contribution up into an academic paper, potentially for FT-DEval.

REFERENCES:

- Åström, T. & Arnold, E., 2023. Meta-evaluation of the Swedish Strategic Innovation Programmes As policy instruments for industrial competitiveness and systems innovation. VR 2023: 16 ed. Stockholm: Vinnova.
- Bergek, A., Hekkert, M. & Jacobsson, S., 2007. Functions in innovation systems: A framework for analysing energy system dynamics and identifying goals for system-building activities by entrepreneurs and policy makers, Gothenburg: RIDE/IMIT, Chalmers.
- Hekkert, M. P. et al., 2007. Functions of innovation systems: A new approach for analysing technological change. Technological Forecasting & Social Change, Volume 74, pp. 413-432.
- Larrue, P., 2021. The Design and Implementation of Mission-Oriented Innovation Policies: A new systemic approach to address societal chalenges, OECD Science, Technology and Industry Papers No 100. Paris: OECD

Tamás Szabó
HÉTFA Research Institute, Hungary

IMPACT EVALUATION OF EU SUBSIDIES FOR R&D ON THE HUNGARIAN FIRMS

ABSTRACT

EU Funds have played a substantial role in Hungarian development policy over the last two decades. In the 2014-2020 programming period, €1.40 billion was allocated directly to research, development, and innovation, representing approximately 6.2% of the total subsidies for that period. Previous research on the 2007-2013 programming period did not differentiate economic development programs based on funding priorities. The main takeaway from those studies is that subsidies have a generally positive impact on performance variables; however, the effect on employment tends to diminish over time.

This research aims to expand upon previous studies in two significant ways. First, as the 2014-2020 programming period has nearly concluded, with most subsidies allocated and projects completed, it is now possible to robustly quantify the effects using more sophisticated econometric methods. Second, previous studies rarely categorized grant programs based on the specific goals of policymakers, such as employment versus capital investment. To align with the scope of the conference, I will focus exclusively on research and innovation subsidies. It is hypothesized that the effects of these subsidies may differ from those targeting other goals.

At the HÉTFA Research Institute, several similar studies have been conducted with nearly identical methodologies but varied research questions. We have the necessary data to conduct this research, including a comprehensive database of balance sheets and profit and loss accounts for Hungarian companies since the late 2000s, which is ideal for such analysis. Additionally, we possess the project database for the 2014-2020 programming period for all operational programs. These datasets can be easily linked using the companies' unique tax numbers. I will calculate the impact of grants on several performance variables, including employment, revenue, tangible asset inventory, and liabilities. Methodologically,

I will combine propensity score matching (PSM) and differences-in-differences (DiD) approaches. This combination will allow me to mitigate measurable differences between the treated and control groups and reduce bias caused by unobserved variables.

Initially, I will calculate the probability of receiving the grant for each observation by estimating the following regression model:

$$p(X) = \Pr(T = 1 \mid X) = f(\alpha + \beta X')$$

In this equation, f(.) = exp(.)/[1 + exp(.)] will denote the logistic distribution function, where T = 1 when the company receives the grant (i.e., treated) and T = 0 when it belongs to the control group (i.e., does not receive the grant). The vector X will contain the values of variables used in the regression, assumed to affect both participation in the treatment and the outcome variables. These variables include various profitability indicators (ROE and ROA), regional and sectoral dummy variables, and other balance sheet items. For each treated company, I will match a control company that most closely resembles it based on the estimated propensity score. As previously mentioned, critically low sample sizes hinder effective algorithmic matching. To address this issue, I will manually pair companies each year based on the estimated propensity score, ensuring:

- The treated and control companies are within a defined proximity threshold based on the score,
- A control company may have multiple treated counterparts,
- Outcome variables are available for the years of the project's contractual start and end,
- Treated cases without a suitable match will be excluded from the study.

Finally, I will conduct a differences-in-differences estimation on this matched treated-control sample. This method compares the outcomes of the treated and control groups at two points in time, before and after the program, and estimates the program's effect by subtracting the two differences. The formal regression equation to be estimated is:

$$lnY_{it} = \beta_0 + \beta_1 T_i t + \beta_2 t + \beta_3 T_i + \beta_4 Z_i + \varepsilon_{it}$$

In this equation, the estimated coefficient of the interaction between time (t = [0,1]) and the treatment indicator (T = [0,1]), $\beta 1$, will indicate the effect size. The vector Z will include additional factors considered to possibly affect the outcome, such as profitability indicators, regional and sectoral dummies, and other balance sheet items.

Additionally, for the examined variables, I will estimate not only the levels of the outcome variables but also their changes (log differences). This approach inherently filters out the effects of unobservable characteristics, thus simplifying the estimation equation to:

$$dlnYi = \beta 0 + \beta 1 Ti + \beta 2 Zi + \varepsilon i$$

where again, $\beta 1$ will indicate the treatment effect. In the presentation, the results will be presented using standard DiD visualization techniques, avoiding extensive tables of coefficients. This approach will include graphical representations such as histograms, line graphs, and scatter plots, which will provide a clear and intuitive understanding of the findings. The outline of my research will be the following:

1. Introduction
    a. Background
        i. Why it is important to support businesses with grants
        ii. What other researchers found
    b. Research objectives
        i. To asses to impact of research and innovation grants on various performance metrics
        ii. To implement a robust methodological approach combining PSM and DiD on a firm level database
    c. Significance of the study
        i. Contribution to policy-making
        ii. Enhancement of evaluation techniques for business support programs
2. Literature review
    a. Theoretical Framework
        i. Economic theories related to subsidies and grants
        ii. The expected impact of financial support on business performance
    b. Empirical Studies
        i. Review of studies using PSM and DiD in similar contexts
        ii. Key findings and gaps in existing literature
    c. Methodological Insights
        i. Strengths and limitations of PSM and DiD
        ii. Justification for combining these methods
3. Methodology
    a. Data Collection
        i. Description of data sources
        ii. Data cleaning and preparation process
        iii. Variables selected for the study
    b. Propensity Score Matching (PSM)
        i. Estimation of the propensity score
        ii. Logistic regression model details
        iii. Selection of covariates (profitability indicators, regional and sectoral dummy variables, etc.)
    c. Matching process
        i. Nearest-neighbor matching
        ii. Criteria for matching and handling of unmatched cases
    d. Differences-in-Differences (DiD)
        i. Explanation of the DiD approach
        ii. Specification of the regression model
        iii. Interaction terms and their interpretation
    e. Combined Methodological Approach
        i. Integration of PSM and DiD
        ii. Advantages of the combined approach

iii. Steps taken to ensure robustness of results

4. Data Analysis and Results
    a. Descriptive Statistics
        i. Summary statistics of key variables before and after matching
        ii. Visualization of propensity score distributions
5. PSM Results
        i. Evaluation of matching quality
        ii. Balance diagnostics (e.g., standardized mean differences)
6. DiD Estimation
    a. Regression results for the combined PSM-DiD model
    b. Interpretation of coefficients and statistical significance
7. Robustness Checks
    a. Sensitivity analysis
    b. Alternative matching algorithms and model specifications
8. Presentation of Results
    a. Graphical representation (histograms, line graphs, scatter plots)
    b. Tabular presentation of regression results
9. Discussion
    a. Key Findings
        i. Summary of main results
        ii. Comparison with previous studies
    b. Implications for Policy and Practice
        i. Recommendations for policymakers
        ii. Practical implications for businesses
    c. Limitations and Future Research
        i. Study limitations
        ii. Suggestions for future research directions
10. Conclusion
    a. Recap of Objectives and Findings
        i. Brief summary of the research objectives and main findings
    b. Final Thoughts
        i. Overall contribution of the study
        ii. Final remarks on the importance of evaluating grant programs
11. References
    a. Complete list of references cited in the study
12. Appendices
    a. Supplementary Tables and Figures
        i. Additional data visualizations
        ii. Detailed regression results

Lisa Neusel, Simon Hirzel
Fraunhofer ISI, Germany

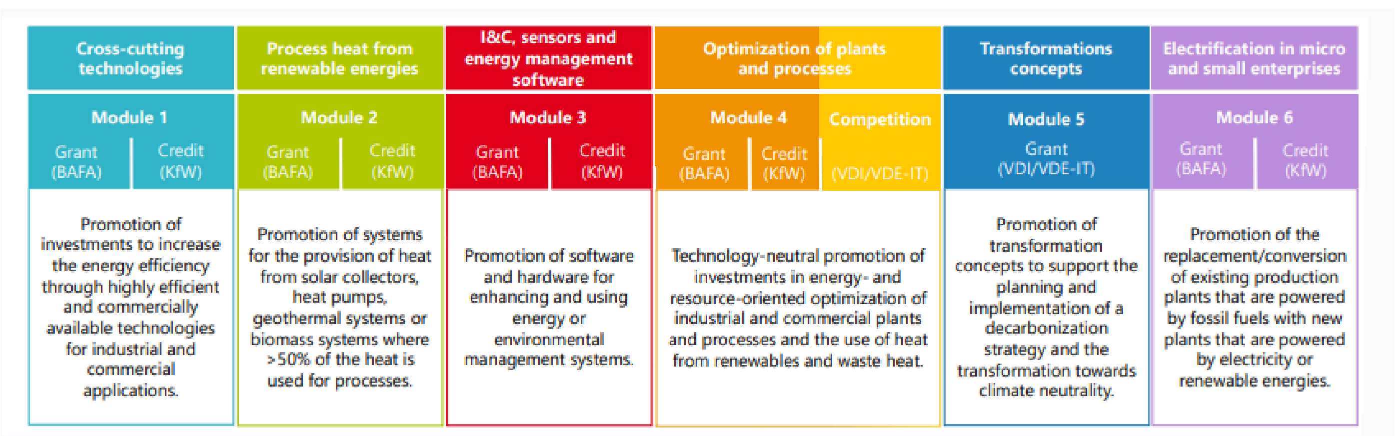
ENERGY AND RESOURCE EFFICIENCY IN THE ECONOMY: THE EVALUATION OF GERMANY'S LARGEST INDUSTRIAL FUNDING PROGRAMME USING MIXED-METHODS

ABSTRACT

Energy use is a very substantial source for anthropogenic greenhouse gas (GHG) emissions. In consequence, improving the deployment of renewable energies and energy efficiency measures is crucial to limit global warming to the 1.5˚C climate target of the Paris agreement. As a response to this, energy policy making uses a variety of regulatory, informational and financial measures to enhance the uptake of corresponding action. The latter in the form of publicly funded subsidy schemes for enhancing the uptake of energy efficiency in companies are an important cornerstone in various countries.

In Germany, a central scheme is the Federal Funding Scheme for Energy and Resource Efficiency in the Economy (EEE). It aims to specifically promote energy-efficient technologies and processes available on the market to support companies in improving energy efficiency. This multi-measure scheme is structured into six modules and offers grant-based, credit-based and competition-based subsidies (Figure 1). According to its latest amendment, the EEE aims at facilitating the implementation of 62,000 measures in companies, thereby targeting saving of 19 TWh of final energy and 7.5 million tons of carbon dioxide emissions from 2022 until the end of 2026. In 2023, the EEE exceeded a volume of 1 billion Euros for the first time.

FIGURE 1: OVERVIEW OF THE ARCHITECTURE OF THE EEE (SOURCE: NEUSEL ET AL. 2024).



Since such schemes as the EEE spend public money, ex-post evaluations are regularly required to review their efficiency and effectiveness. Also, reporting requirements on measures addressing European and national energy efficiency and climate targets have increased considerably in recent years. On the European level, the most detailed ones are requirements for the communication of measures and methods for the implementation of Article 8 of the Energy Efficiency Directive recast (Directive 2023/1791/EU, Annex V). Germany's energy and climate targets, with the overall goal of achieving GHG neutrality in 2045, are legally defined in the revised Federal Climate Change Act (KSG) of 2021 and the Energy Efficiency Act (EnEfG) of 2023. Both include reporting requirements in several places, which also include the evaluation of measures in a quantitative manner. Against this background, this contribution aims to illustrate such a quantitative evaluation in a twofold way: It will outline the mixed-method approach underlying the evaluation and it will yield an overview of the most recent evaluation results of the EEE.

The evaluation is based on a methodological framework formalized in Schlomann and Voswinkel et al. (2020) within the EEE's predecessor programme (Hirzel and Antoni et al. (2019)), as well as on previous expertise in energy policy evaluation. The purpose of this methodology is to:

- Monitor target achievement: To what extent were the objectives of the funding achieved?
- Assess impact: Is the funding the cause of the impact or suitable for triggering it?
- Control efficiency: Are both the funding provided (efficiency of implementation) and the objectives achieved (efficiency of measures) in an economical manner?

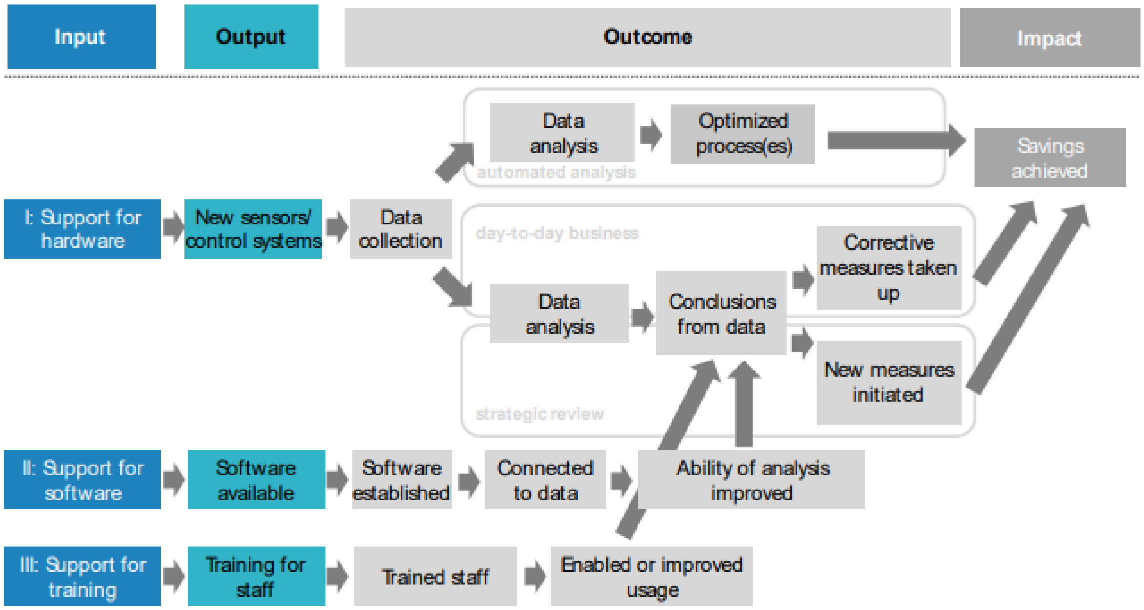
TABLE 1: OVERVIEW OF THE EVALUATION METHODOLOGY (SOURCE: HIRZEL AND SCHLOMANN 2022).

Step and purpose	Main tasks
1: Characterization Description of the covered policy measures	• General **outline of the measure** covering its type, target group/sectors, budget, funding bodies/implementation agencies, legal basis, related policy measures and the funding process • Analysis of the **impact model** of the measure • Consideration of potential **distortions** (e.g. overlaps, double counting, side-effects such as free-rider effects, spill-overs or follow-up-effects)
2: Framework data Definition of common data and assumptions	• Definition of **harmonized input data** (e.g. emissions factors, primary energy converters) • Provision of **defaults choice lists** (e.g. lifetimes by type, energy prices)
3: Targets and requirements Identification of the targets of the overall programme, its individual policy measures and of the specific requirements to the evaluation	• Description of **requirements and expectations** to the evaluation • Analysis of **top-down targets** for energy efficiency improvements based on governmental documents, directives and laws • Analysis of **bottom-up targets** of individual support schemes from ex-ante estimation, funding guidelines • Definition of the **main areas of interest** for the evaluation
4: Indicators Setting up performance values to measure the achievement of targets	• Selection of **indicators that reflect progress** in the areas of interest • **Operationalization** of the indicators: choice between qualitative/quantitative type, description and delimitation, computational model, type of result, units (quantitative) or scales including interpretation rules (qualitative)
5: Data collection Identification and collection of data for establishing the indicators	• Establishing a **data collection concept** based on the selection and setup of the indicators • Implementation of the **data collection process**
6: Data review Processing incomplete or missing information	• **Review of data** (e.g. error correction, missing parameters) • Method selection and implementation of **backcasting and projections** of data where needed
7: Data analysis Processing of the data to measure the achievement of the target values	• Selection of appropriate **method of analysis** (descriptive/analytical) • Computation of **gross values** for indicators
8: Net impact estimation Eliminations of distortions in the results	• **Identification of distortions** • Computation of undistorted **net values** for indicators • Conclusions for the **individual measures**
9: Overall assessment Merging individual results	• Determination of **areas for aggregation** and comparison • **Correction for double counting** when aggregating quantitative values • Computation of the **overall assessment** • Formulating conclusions for the **entire scheme**

The overall methodology consists of nine steps (Table 1). Selected steps are detailed further in the following to illustrate the methodology.

One of the initial steps which is a part of the characterization of the measures is the development of an impact model (step 1). The impact model is a logical causal chain and deliberate simplification of the influences to make impact relationships manageable in the evaluation. The basic impact model follows an input-output-outcome-impact-logic where the input depicts the effort put into the program, the output reflects the immediate result, the outcome the content-wise changes and the impact the final result of the intervention on the level of the overall aim. For each module of the EEE, a specific impact model is used to investigate the individual impact. Figure 2 illustrates such an impact model for Module 3 in the EEE which is a support program for measuring and control equipment including support for software and training. The impact model depicts the logic of the overall intervention finally resulting in energy savings as an impact.

FIGURE 2: ILLUSTRATION OF AN IMPACT MODEL AT THE EXAMPLE OF MODULE 3 OF THE EEE (SOURCE: NEUSEL AND HIRZEL ET AL. (31.10.2023)).



The impacts such as energy savings but also various other aspects are mostly quantitatively assess along a set of key performance indicators (KPI, step 4), adding up to a total of more than 80 indicators (including sub-indicators) across all categories. The table below provides a first brief overview of the KPI chosen for the evaluation on the level of individual measures. They are structured along the established core evaluation areas, i.e. indicators for reviewing target achievement (A), effectiveness (B) and economic efficiency (C). The indicators of category B are also included as net values in the analyses of categories A and

C. In addition, a group of indicators of general knowledge interest (G) is included. Further additional indicators cover the perception of the procedural implementation (D). Most of the KPIs are quantitative values (e.g. GHG savings in tonnes of CO2-eq.), yet some are provided qualitatively (e.g. funding process quality).

TABLE 2: OVERVIEW OF KEY PERFORMANCE INDICATORS FOR THE EEE

(G) - General knowledge interest: Structural data on applications, approvals and funding
(A) - Target achievement: To what degree have the established targets been achieved?
(B) - Effectiveness: To what degree is the measure causal to the achievements?
(C) - Economic efficiency: How efficient is the measure from the implementer's perspective with regard to achieving the targets and concerning the use of resources?
(D) - Procedural implementation: How is the operational implementation perceived?
(E) - Specific knowledge interest: Module-specific questions that go beyond the specified evaluation objectives

In particular the quantitative KPI are determined using two sources of information (step 5): Data from the administration of the support scheme and data from the beneficiaries themselves. The administrative data contains information submitted during the application and includes information on the beneficiary (e.g. name, regional allocation, company size) and various financial and related information on the activity submitted for funding. Data from the participants is collected via a survey among beneficiaries. In the most recent iteration, nearly 11,000 beneficiaries were invited and the typical response rate was about 20% in former years. The survey itself consists of various common questions across all modules and specific questions addressing particular aspects related of the individual modules or implementing agencies. It typical duration is in the order of magnitude of 15 to 25 minutes.

Using this data, gross KPI values are determined. To analyse for cases where the EEE might not have been the only or a very relevant reason for the investments in energy efficiency, an effect adjustment is carried out (step 8). For this, the questionnaire contains several control questions, related to the extent and role of the funding scheme for the activity. Based on an effect adjustment with several steps, net impact values are estimated which seek to both take negative effects (e.g. free-riders) and positive effects (e.g. spill-overs) into account.

TABLE 3: APPROACH FOR NET IMPACT ESTIMATION (SOURCE: VOSWINKEL (2019)).

Impact / Effects	**Description**
Gross value	Impact before considering effects
- Free-rider effects	Saving that would have occurred without policy
+ Spill-over effects	Effects on third parties and other areas not directly credited to the programme
+ Follow-on effects	Effects through not yet completely realized actions
- Double-counting	Effects through interaction of measures
= Net value	Impact after adjusting for effects

Expected results of the final submission The dynamic character of the EEE, reflected by its various amendments and extensions over the years, repeatedly poses new methodological and operational challenges to the annual evaluation. The EEE is evaluated on an annual basis and the evaluation activities for the year 2023 are currently taking place. Besides providing an overview of the most recent results and the general methodology, the final contribution seeks to point out some of the more recent methodological challenges to the EEE yet not covered in prior publications. Among others, this will include methodological modifications such as the including of new emission factors, among others the introduction of dynamic values for electricity to take the expected decarbonisation of the electricity sector further into account. In addition, the recently introduced modules 5 on transformation concepts and 6 on the electrification in micro and small companies are evaluated for the first time and require changes to the prior established quantitative approaches for evaluating their impact. The discussion of these changes can help to prepare similar quantitative evaluations of (other) support schemes in the energy domain.

REFERENCES:

- Hirzel, S., O. Antoni, S. Heinrich, D. Jessing, A. Paar, M. Pehnt, B. Schlomann and F. Voswinkel (2019). Evaluierung und Weiterentwicklung des Energieeffizienzfonds (Projekt Nr. 63/15). Karlsruhe/Heidelberg/Basel/W¸rzburg.
- Hirzel, S. and B. Schlomann (2022). Evaluating Multi-measure Schemes for Enhancing Energy Efficiency: The German Energy Efficiency Fund. Proceedings of the eceee summer study 2022.
- Neusel, L., S. Hirzel, C. Berger, B. Schlomann, S. Heinrich, K. Weinert, A.-M. Grodeke, P. Radgen, A. T. Anzaldo Grundler, K. Schumacher, C. Nissen and S. Ludig (2023). Evaluation der “Bundesförderung für Energie- und Ressourceneffizienz in der Wirtschaft” (Zuschuss und Kredit/Fˆrderwettbewerb). Karlsruhe, Basel, Stuttgart, Berlin.
- Neusel, L., S. Hirzel, K. Weinert; S. Heinrich; A.-M. Grodeke (2024): Efficiency Unleashed: Evolution and Impact of Germany's Funding Scheme for Energy and Resource Efficiency in the Economy. Accepted submission for the eceee summer study 2024.
- Schlomann, B., F. Voswinkel, S. Hirzel, A. Paar, D. Jessing, S. Heinrich, O. Antoni and M. Kahles (2020). Methodikleitfaden f¸r Evaluationen von Energieeffizienzmaflnahmen des BMWi (Projekt Nr. 63/15 - Aufstockung). Karlsruhe/Heidelberg/Basel/W¸rzburg.
- Voswinkel, F. (2019). Gotta catch ‘em all - catches to evaluating heterogenous energy efficiency programmes, eceee.

Klaus Schuch, Olga Bolibok
ZSI – Centre for Social Innovation, Austria

FP10 - A FIELD OF TENSION BETWEEN OPENNESS TO RESEARCH AND STRATEGIC DIRECTIONALITY

ABSTRACT

The debate on the implementation of Horizon Europe (HEU) in its remaining term and initial considerations on the organisation of a tenth EU FP (FP10) are currently equally high on the political agenda and in the discourse of the stakeholder groups involved in R&I. This is partly due to the fact that HEU is currently being scrutinised as part of the interim evaluation and, at the same time, an initial proposal from the European Commission for the design of the next EU FP has been announced for 2025. At the beginning of December 2022, the European Commission conducted a public consultation on the past, present and future of the EU FPs, some of the results of which are being incorporated into the interim evaluation of HEU (see EC DG RTD, 2023b). At the end of 2023, the European Commission's Directorate-General for Research and Innovation (DG RTD) also appointed a high-level expert group to accompany and finalise the interim evaluation of HEU until October 2024. Although the official negotiations for FP10 will not begin before mid-2025, the course is already being set and the member states have initiated an initial consultation process under the umbrella of ERAC together with DG RTD.

Against this backdrop, the German Federal Ministry of Education and Research (BMBF) commissioned a study, which was carried out by Joanneum Research and the ZSI - Centre for Social Innovation between April 2023 and March 2024. One of the aims of the study was to explore what German stakeholders in the R&I sector expect from the next Framework Programme, FP10. The overall ex-ante results of this study served as an information basis for the BMBF to build up its negotiating position with regard to the design and organisation of FP10.

The purpose of this paper is to subject key structural elements discussed ex-ante in relation to the next Framework Programme to a discourse analysis based on the assessments of German experts from the R&I sector. In this paper, we focus on the perceived tension between (strategic) directionality and open (non-directional) research and the implications this could have for the structure of FP10.

To identify the expectations of German R&I experts as regards FP10, 28 interviews were conducted with them, which were validated in a subsequent reflection workshop with a further six international experts. The summarised results were then fed into a so-called "future dialogue", which was held at the BMBF in Bonn in the presence of around 30 participants (mainly from R&I policy-making and policy-delivery). In addition, a questionnaire on the ex-ante expectations with regard to the 10th Framework Programme

was sent out, which was answered by approx. 150 (depending on the question) funding recipients from H2020 or HEU who had received funding more than once, in order to get a rough initial overview of the expectations for FP10. The results of the various methods were triangulated in terms of content.

The expert interviews followed a pre-structured discussion guide and focussed on the following topics:

- Significance of open strategic autonomy and technological sovereignty incl. defence research and critical RTI infrastructures in FP10
- Importance of international R&I cooperation between value debate, research security and global challenges
- Support for less innovative European countries in FP10
- Balance between basic research, applied research and innovation in FP10
- Significance of directional and thematic research (EU missions, EU partnerships) in FP10
- Complementarity between the EU-FP and national and other European programmes

The interviewees included high-level representatives of German universities, non-university research institutions, policy-makers, umbrella organisations and agencies as well as representatives from the corporate sector who are very familiar with the subject. The interviews lasted approximately one hour. The transcribed interviews totalled just under 800 pages. They were analysed with the help of maxQda.

As already mentioned, our focus in this presentation is on the perceived tension between (strategic) directionality and open (non-directional) research and the implications this could have for the structure of the next European R&I Framework Programme. In view of budget cuts and reallocations that have already taken place in HEU as well as emerging areas of tension between R&I policy, transformation and industrial policy (cf. Juh·sz, Land and Rodrik, 2023) as well as the pursuit of technological sovereignty (cf. Edler et al., 2020) and overcoming global challenges and the associated dealings with countries such as China and any resulting lines of conflict, the discussion on the design of the tenth EU FP promised to be a lively one – and it was. The opinions of the experts interviewed, and their corresponding arguments are analysed along the following lines of conflict, which are then elaborated in more detail in the presentation and the paper:

- Strategic considerations on "open autonomy", strengthening European resilience and technological sovereignty are generally favoured, but what these considerations mean for FP10 in practice is controversial. There is, however, a joint understanding that FP10 should serve the goal of scientific and technological sovereignty, in particular through anticipatory research on key technologies.

- At the same time, it is emphasised that FP10 must not be overloaded with agendas that are more of an industrial policy reform nature and do not genuinely fall within the R&I area (e.g. policies to make European locations more attractive; protection measures for strategic areas; economic upscaling; strategic public procurement, etc.). Instead, such agendas would require a separate European industrial reform programme, which should have an interface with FP10, but complement it.
- These tensions can be illustrated by the example of defence research. Russia's attack on Ukraine has created a "window of opportunity" for European support for defence research, which is also ideally supported by many German R&I experts. At the same time, opinions differ widely as to whether FP10 is a suitable framework for defence research. There are fears that, firstly, instruments are required that do not fit the FP (strong focus on procurement), secondly, the budget requirements would put the rest of the FP under severe pressure, thirdly, there will be latent tensions in consortia with regard to open science and, fourthly, some universities and research institutions that have dedicated themselves exclusively to civilian research would come under internal pressure. At the same time, however, it is recognised that defence research is generally cutting-edge research and that spill-overs into civilian applications would be important.
- Of course, the opinions of the experts interviewed also depend on the institutional interests they represent. It is therefore not surprising that experts from the academic environment were strongly in favour of securing the ERC, while the experts from industrial research in particular emphasised the importance of Pillar 2. It is interesting to note, however, that both groups warn against an increasing fragmentation of the framework programme, although they are generally in favour of retaining the 3-pillar structure. They emphasise the need to overcome the perceived compartmentalisation through more collaborative research, including at lower TRLs (both in the ERC and in Pillar 2).
- Pillar 3, on the other hand, is perceived as a "start-up" pillar and is accepted as such as long as it remains limited in terms of budget compared to the other two pillars. It is not the focus of either large-scale industry or universities and non-university research institutions. The EIT, which is anchored in Pillar 3 of Horizon 2020, is viewed most critically. It is perceived as rather non-transparent, bureaucratic, geared towards TRLs that are too high and with limited efficiency and efficacy.
- The relevance of EU missions is currently widely recognised, although many players in the innovation ecosystem are not yet able to position themselves within the missions or clearly recognise the interplay between EU missions, clusters or activities in other pillars. There is also a clear call for internal alignment and harmonisation of the instrument portfolio of European programmes, in particular the coordination between R&I and sectoral directoratesgeneral, agendas and programmes. Some see the missions as a cross-European structural

element that should reach beyond the EU FP and provide a structural framework for the integration of other European funds (such as ERDF, EAFRD, ESF, etc.). Most experts also supported the idea to further mirror the EU missions at national level. The future of missions in RP10 is controversial. For some, they could even replace the clusters, while others see them as being better off outside RP10. But these are only extreme opinions. Most are in favour of a cautious middle way, if their position in Pillar 2 is clarified, as this is currently not understood, and secondly if the research content is more substantiated (also with transdisciplinary research and SSH).

- The European Partnerships are viewed in a critical light too, with almost all interviewees admitting that they know too little about them, which is an interesting observation in itself. The public-public partnerships are seen by policy-makers as an important structural element. The industrially orientated public-private partnerships appear relevant to most of the experts. However, industry representatives also described them as bureaucratic, too complex and cost-intensive. The creation of strategic research agendas is appreciated. For FP10, they should be dimensioned in terms of their number (less) and size (less).
- The future of international R&I cooperation in FP10 is clearly caught between the poles of directionality and openness, with particular relevance for the discussion on technology leadership and open strategic autonomy. All interviewees stressed the importance of international R&I cooperation, in particular to tackle global challenges jointly. The business community also argues that international R&I cooperation, which is also to be made possible in FP10, offers opportunities for market development. Furthermore, it was clearly stated that international cooperation could also help to ensure the resilience of European R&I. The empirical evidence also shows that both the opportunities for European internal cooperation and those with third countries are an important determinant for supporting young researchers and strengthening human research capacities. The debate on values currently being conducted at the R&I policy level with regard to responsive international R&I cooperation, particularly in relation to countries such as China, is perceived as important, but should not ultimately make international cooperation impossible. Rather, differentiated implementation modalities for responsible international R&I cooperation should be created, which may speak in favour of a new INCO programme.

During the conference, the implications of the tensions between strategic directionality and open (non-directional) research for the structure of the next framework programme will be presented.

REFERENCES:

- EC DG RTD (2017): Interim Evaluation of Horizon 2020, Commission staff working document, URL: https://op.europa.eu/en/publication-detail/-/publication/33dc9472-d8c9-11e8-afb3- 01aa75ed71a1/language-en, letzter Zugriff: 18. M‰rz 2024.
- EC DG RTD (2021): Horizon Europe, Strategic Plan 2021-2024, URL: https://www.eeas.europa.eu/sites/default/files/horizon_europe_strategic_plan_2021-2024.pdf, letzter Zugriff: 18. M‰rz 2024.
- EC DG RTD (2022): European Partnerships in Horizon Europe infographics, URL: https://researchand-innovation.ec.europa.eu/knowledge-publications-tools-and-data/publications/allpublications/european-partnerships-horizon-europe-infographics_en, letzter Zugriff: 22. Mai 2024.
- EC DG RTD (2023a): Horizon Europe, Strategic Plan 2025-2027, URL: file:///C:/Users/Marlies.Schuetz/Downloads/horizon%20europe%20strategic%20plan%202025- 2027%20analysis-KI0423276ENN.pdf, letzter Zugriff: 18. M‰rz 2024.
- EC DG RTD (2023b): Looking into the R&I future priorities 2025-2027, Synopsis Report, URL: https://op.europa.eu/en/publication-detail/-/publication/e8f722ec-dff7-11ed-a05c01aa75ed71a1/language-en/format-PDF/source-284554549, letzter Zugriff: 18. M‰rz 2024.
- EC DG RTD (2023c): Study supporting the assessment of EU Missions and the review of mission areas – Mission areas review report, URL: file:///C:/Users/Marlies.Schuetz/Downloads/study%20supporting%20the%20assessment%20of%20eu%2 0missions%20and-KI0423786ENN-2.pdf, letzter Zugriff: 22. Mai 2024.
- EC (2023d): Horizon Europe implementation, Key data for 2021-2022, URL: https://op.europa.eu/en/publication-detail/-/publication/66cb5a2c-0b31-11ee-b12e01aa75ed71a1/language-en/format-PDF/source-287848713, letzter Zugriff: 21. Mai 2024.
- EC (2024a): Bericht der Kommission an das Europ‰ische Parlament und den Rat, Ex-post-Bewertung des EU-Rahmenprogramms f¸r Forschung und Innovation Horizont 2020, COM(2024) 49 final, URL: https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:52024DC0049, letzter Zugriff: 21. Mai 2024.
- EC (2024b): Weiflbuch ¸ber Optionen f¸r eine verst‰rkte Unterst¸tzung von Forschung und Entwicklung zu Technologien mit potenziell doppeltem Verwendungszweck, URL: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/14060-FuE-zu-Technologienmit-doppeltem-Verwendungszweck-Optionen-fur-die-Unterstutzung_de, letzter Zugriff: 22. Mai 2024.
- ECFR (2019): Empowering EU Member States with strategic sovereignty, URL: https://ecfr.eu/wpcontent/uploads/1_Empowering_EU_member_states_with_strategic_sovereignty.pdf, letzter Zugriff: 22. Mai 2024.

- ECIPE (2020): Europas Streben nach Technologiesouver‰nit‰t, Chancen und Risiken f¸r Deutschland und die Europ‰ische Union, ECIPE Occasional Paper 05/2020, URL: https://ecipe.org/wpcontent/uploads/2020/07/ECI_20_OccPaper_05-2020_LY03.pdf, letzter Zugriff: 22. Mai 2024.
- Edler, J. et al. (2020). Technologiesouver‰nit‰t – Von der Forderung zum Konzept. Fraunhofer ISI, Karlsruhe, Juli 2020.
- Hofmann, K., Janger, J. und F. Unterlass (2023). Technologische Souver‰nit‰t. Empirische Bestimmung und FTI-politische Implikationen. ÷sterreichisches Institut f¸r Wissenschaftsforschung, Wien, Juni 2023.
- Juhasz, R.; Lane, N. und D. Rodrik. (2023): The New Economics of Industrial Policy, NBER Working Paper Series, URL: https://www.nber.org/papers/w31538, letzter Zugriff: 18. M‰rz 2024.
- Science Europe (2024): Science Europe's Initial Position: European Commission's Public Consultation on the White Paper on R&D on Dual-Use Technologies, URL: https://scienceeurope.org/ourresources/science-europe-s-initial-position-european-commission-s-public-consultation-on-the-whitepaper-on-r-d-on-dual-use-technologies/, letzter Zugriff: 22. Mai 2024.
- The Guild (2022): Horizon Europe Application Experiences (survey), URL: https://www.theguild.eu/publications/statements/the-guild_horizon-europe-survey_13-january-2022.pdf, letzter Zugriff: 22. Mai 2024.

Position papers of EU Member States and organisations: https://era.gv.at/fp10/next-frameworkprogramme/position-papers/

Stephanie Francis Grimbert, Nicholas Vonortas
University of Deusto, Spain
University of Campinas, Brazil

AN ATTENTION-AUGMENTED REAL OPTIONS APPROACH TO PUBLIC R&D INVESTMENTS

ABSTRACT

A real option is “the right, but not the obligation, to take an action in the future” (Amram & Kulatilaka, 1999 p.5). In other words, option creation may or may not lead to the exercise of that real option regarding, for instance, innovation projects (Klingebiel & Rammer, 2021; Ross, et al., 2018), corporate venturing (Ceccagnoli, et al., 2018; Tong & Li, 2011), or new business development (Klingebiel, 2012; McGrath, 1999). By providing the opportunity to adapt flexibly to new information, real options enhance the economic value of an investment by limiting the downside losses expected initially (Kogut & Kulatilaka, 1994; Trigeorgis, 1997). The real options approach has gained momentum in the field of strategic management (Li et al., 2007; Trigeorgis & Reuer, 2017), on the one hand as strategic reasoning exempt of analytical modelling (real options reasoning) and on the other as a formal modelling method simulating real option valuation (real options valuation). Real options theory contributes to informing heterogeneity between organizations by “identifying critical bi-directional linkages” (Trigeorgis & Reuer, 2017. p.52) in such a manner that real options both emerge from and enhance heterogeneity when successfully recognized and managed (Tong & Reuer, 2007). The management of real options therefore causes variance in the value extracted from them (Bowman and Hurry, 1993; Kogut and Kulatilaka, 1994; McGrath, 1999; Tong & Reuer, 2006, 2007; Trigeorgis & Reuer, 2017), which places a high demand on managerial capabilities for execution.

Derived from the valuation of financial options (Black & Scholes, 1973; Myers, 1977), the conventional view on the value of real options considers that it is an increasing function of the uncertainty about the future value of the underlying asset by providing managerial flexibility as opposed to having to commit to an investment in the face of uncertainty. However, bearing in mind that “one key challenge for the formal modeling of real options, compared to basic financial options, is that multiple sources of uncertainty can affect the value of many real options” (Trigeorgis & Reuer, 2017 p.45), recent developments in real options theory have attempted to model the impact of different types of uncertainty on the value of real options. Posen and colleagues (2018) discriminate “prospective” (i.e., future) and “contemporaneous” (i.e., current) categories of uncertainty, the latter being introduced to “relax the assumption that firms have objective information about the asset value that is both accurate and precise at any point in time” (Posen et al., 2018 p.1118). As a result of the value of real options being subject to both prospective and contemporaneous uncertainty,

real options theory embraces a behavioral turn. As advocated by Trigeorgis and Reuer (2017), the inclusion of behavioral perspectives to real options theory enables exploring the effects of bounded rationality to the reliance on real options, according to which the value of real options would tend to be overestimated (Posen et al., 2018; Smit & Kil, 2017) because of potential execution errors.

The potential of real options theory has been explored in the literature on R&D investments (e.g., Perlitz et al., 1999; McGrath & Nerkar, 2004). Considering R&D investments in real option terms provides an analytical grid for perceiving the increased financial value of a firm's innovative activities. Using this real option lens, new venturing projects are seen as small investments delimiting a learning space aimed at gradually improving the firm's knowledge about the future potential of a large panel of different technologies (Vanhaverbeke et al., 2008; Trigeorgis & Reuer, 2017; Ross et al., 2018). But given that "a wealth of information creates a poverty of attention" (Simon, 1971 p.40), approaching R&D investments in real options terms also allows for recognizing the attentional constraints which affect the realization of the potential value of real options (Barnett, 2005) by imposing the allocation of attention across multiple projects (Kim et al., 2016). Behavioral real options theory usefully complements an attention-augmented perspective to the valuation of real options in the context of R&D investments. The volatility associated with R&D projects is determined by market uncertainty and technical uncertainty, which induce complexity for the accurate transfer of financial option pricing techniques to actual investment decisions (Vonortas & Desai, 2007). Hence, "there is an option value of additional information" (Huchzermeier & Loch, 2001 p.99) about project progress and market characteristics which explains "the gap between the financial payoff variability [...] and operational uncertainty" (Ibid. p.86) and requires leaving space for improvement as a complementary type of real option under operational uncertainty.

Staged public R&D investments as is the case in the Small Business Innovation Research program in the US and its equivalent in Europe, the Pre-Commercial Procurement scheme, have also benefited from real options valuation techniques. They are considered more appropriate to fully capture the strategic value of investment opportunities than traditional capital budgeting techniques, which tend to penalize long-term investments (Vonortas & Hertzfeld, 1998; Vonortas & Lackey, 2003; Vonortas & Desai, 2007; Belz & Giga, 2018) endowed with "a high-risk/high-rewards approach to meet the technology vision" (Bonvillian, 2014 p.7). However, the literature on the strategic management of R&D investments through an attention-augmented real options approach has yet to be adapted to the specific context of public agencies, despite the particular attentional constraints exerted by relying on externally distributed intelligence. Public R&D investments assist the private sector in speeding up the introduction and commercialization of innovations endowed with strong asset-specificity (Casady et al., 2023), but are aimed at increasing service performance and public value, the utility function of which is unspecific. We address this gap by applying to public R&D investments an attention-augmented real options approach, which we define as the right but not the obligation to acquire an asset given attentional constraints when there is both prospective and contemporaneous uncertainty. Our framework sheds light on the bi-directional linkages between an attention-augmented real options approach and the heterogeneity of public agencies. Our first research question examines the effects of public sector heterogeneity in terms of an attention-augmented real options approach. Our second research question looks at how an attention-augmented real options approach can inform the strategic management of public sector heterogeneity.

Jon Mikel Zabala Iturriagagoitia, Stephanie Francis Grimbert, Juan Carlos Salazar Elena
University of Deusto, Spain
University of Deusto, Spain
Autonomous University of Madrid, Spain

MEASURING PUBLIC PROCUREMENT FOR INNOVATION IN EUROPE

ABSTRACT

Public procurement is a key public policy instrument. Existing estimates suggest that 19-20% of a country's GDP is produced through government procurement (Kahlenborn et al., 2010). Spending on public procurement is therefore many times higher than spending on R&D, which is considered to be the main vehicle for fostering innovation.

Procurement strategies can be structured in a 'regular' way, involving the procurement of products already available on the market to meet specific needs, a method colloquially referred to as "off-theshelf" (Edquist et al., 2015). However, in cases where there are no off-the-shelf solutions available on the market the responsibility of procurement lies in driving innovative solutions to address recognized problems. Effectively taking on this challenge requires the precise definition of problems, their translation into requirements and the subsequent communication of the latter to the market. In this sense, public procurement for innovation (PPI) emerges as a key policy instrument designed to articulate the demand for innovations, fostering their widespread adoption and diffusion (Edquist and Zabala-Iturriagagoitia, 2012).

PPI is a strategic process in which a contracting authority formulates and places an order for a good, service or technology that does not currently exist (because the business sector has not yet developed it) but is essential to address specific functions in order to solve societal, departmental, ministerial or related problems. The fundamental objective of PPI is to address unmet public, social and environmental needs or mitigate challenges that are often global in nature but may have national or local dimensions simultaneously (Saussier and Tirole, 2015). It is imperative that this procurement approach leads to innovative solutions before delivery to the contracting authority occurs. Therefore, innovations derived from the PPI must offer a substantial improvement compared to existing products or services at the time of bidding. PPI can accelerate technological progress, as well as the adoption and diffusion of innovations, which could reshape the overall industrial landscape (Bleda and Chicot, 2020). PPI does not only yield substantial social benefits, but it also generates tangible economic advantages for both suppliers and their corresponding supply chains, increasing their overall socioeconomic contribution (Wesseling and Edquist, 2018).

The extant literature provides evidence that public procurement proves to be a more

effective instrument in fostering innovation and entrepreneurship than other traditional instruments such as R&D subsidies. However, despite its potential, until now it has not been possible to carry out a rigorous systematic evaluation of public procurement practices due to the lack of adequate data and comparative methodologies that would allow a comprehensive analysis of the degree of use of this instrument. As a result, there is an absence of studies with empirical analyses that offer a comprehensive measurement of this central policy instrument, beyond individual case studies.

The aim of this paper is to offer a systematic evaluation of the use of PPI in Europe, both at national and regional levels. It is based on the information provided by the Tender Electronics Daily (TED) database of the European Commission, for years 2016-2023. In order to identify those tenders that have an innovative component, natural language processing and machine learning methodologies will be used. On the one hand, natural language processing (Just, 2024) allows us to identify patterns when faced with large amounts of information, as in the case of public procurement tenders, where all contracts managed by national and regional contracting authorities above the European thersholds are included, representing millions of procurement contracts. The use of the previous methodology allows us to identify:

- the tenders that incorporate functional requirements (i.e. the identified needs have been translated into functional terms); •
- the award criteria identified in the tenders, and the relative weight (i.e. %) of each of them in the evaluation of the proposals (i.e. the literature identifies that in PPI projects price is not the main award criterion, but that a wide variety of other criteria are used);
- the extent to which the procurement of key products is currently being undertaken in Europe (e.g., those products and technologies included in the EC guidance document on the implementation of STEP - published on May 12th 2024 that are critical for European security and autonomy)

The previous methodology allows us to offer a map at national and regional levels on: (i) the % of PPI tenders over total tenders; (ii) the budget that PPI tenders represent (in %) over total tenders and over the GDP(%); (iii) which are the most used procedures in PPI tenders of the total procedures identified in the European directives on public procurement (i. e. open procedure, restricted procedure, competitive dialogue, competitive dialogue with negotiation, partnership for innovation); (iv) the % of projects not categorized as PPI and that could be characterized as such. e. open procedure, restricted procedure, competitive dialogue, competitive dialogue with negotiation, innovation partnership); and (iv) the products and technologies in which Europe shows a situation of critical dependency, and for which a lack of procurement contracts is being observed.

Angela Wroblewski
Institute for Advanced Studies (IHS), Austria

ERA ACTION 5 AND ITS MONITORING – INSTRUMENTS TO STRENGTHEN GENDER EQUALITY IN THE EUROPEAN RESEARCH AREA?

ABSTRACT

Gender equality has been one of the main objectives of the European Research Area (ERA) for about two decades, although this objective is currently being strongly challenged by anti-gender movements and conservative currents. In the current ERA period (2022-2025), Member States have committed themselves to a selected set of actions to advance the ERA. One of these actions focuses on gender equality and inclusiveness (ERA action 5). Already in the ERA period 2016–2020, the ERA roadmap formulated gender quality as a priority.

A comparison of both strategies to foster gender equality in research and innovation shows further developments as well as setbacks. Further developments referred to experiences with the implementation of the ERA roadmap 2016-2020 and recommendations formulated based on its monitoring (e.g. regarding the conceptualization of gender equality). Other recommendations have not been considered in the design of ERA action 5. These refer among others to the monitoring of policy implementation and the use of the monitoring for a policy discourse.

Wroblewski (2021) argued for a strong monitoring system to strengthen gender equality in ERA. The criticism primarily related to the quantitative focus of the monitoring. It has been shown that the proportion of women in Grade A is not a meaningful indicator of gender equality, and that qualitative information should be used to monitor policy implementation. For the period 2016-2020, progress towards gender equality has been measured by the so-called headline indicator (proportion of women in Grade A). In the current period, the monitoring (ERA dashboard) contains five gender related indicators (Share of women in grade A positions in HEIs, Proportion of papers with mixed gender authorship, Proportion of women in authorship, of the top 10% most cited publications, Women in digital Index, Proportion of women among doctoral graduates by narrow STEM fields). The dashboard is part of the ERA Monitoring Mechanism recommended by the Pact for R&I (Council for the European Union 2021a). The monitoring mechanism should ensure a proper basis for evidenceinformed policy making in the ERA and to support and facilitate the implementation of the ERA Policy Agenda at both European and national levels. According to the Pact, the ERA Monitoring Mechanism should encompass the following elements:

- an ERA Scoreboard, which monitors progress towards the ERA objectives at Union level. The ERA Scoreboard should be updated regularly and should assess the overall consolidation and collective progress of ERA priorities. It should only display aggregated data at Union level.

- a more detailed ERA Dashboard monitoring progress towards the ERA objectives at national level, through a rich combination of relevant input, outcome and impact indicators and qualitative analyses that accommodate the different circumstances of Member States and that relate to the ERA priorities.
- regular policy dialogues between the Member States and the Commission – both bilaterally and multilaterally – to actively assess and guide the implementation of the ERA policy agenda, in particular through the sharing of best practices and mutual learning exercises. The Commission will provide further support through the Horizon Policy Support Facility and the Technical Support Instrument.
- an ERA policy online platform, where the Member States and the Commission should share information on their current and planned policies and programmes that contribute to implementing the ERA Policy Agenda.
- a review of the implementation of the ERA policy agenda by the Commission taking place every 18 months, including a report for consideration by the Council on the state of play of its implementation in view of steering the ongoing ERA Policy Agenda
- an annual report provided by the Commission to each Member State on its progress, in support of the regular policy dialogues between Member States and the Commission.

WP5 of the GENDERACTIONplus project (Horizon Europe) aims at providing a monitoring of ERA action 5 implementation at national level which complements the quantitative approach of the ERA dashboard. The monitoring approach of GENDERACTIONplus (Wroblewski 2023) is based on the assumption that the effective implementation of policies follows a complete policy cycle. Ideally, gender equality policies are based on a baseline assessment of the status quo regarding gender equality. What are the main challenges to be addressed? Which mechanisms produce inequalities?

How could these inequalities be tackled? Based on the results of the gender analysis, gender equality objectives are formulated. These objectives are the starting point for the development of concrete measures. These measures are implemented, monitored and in an ideal world evaluated.

This ideal model can be formulated for the European level as well as for the national or institutional level. Ideally, the levels influence and reinforce each other. This means that the objectives or measures formulated at European level are adopted at national level or, where necessary, adapted to national circumstances. Similarly, at the institutional level, national equality objectives are adopted or gender equality priorities are set according to the national context. In this ideal world, neither national objectives contradict those at EU level, nor institutional objectives contradict those at national level. Any reservations or resistance will be raised and discussed in the appropriate political discourse.

The first report of GENDERACTIONplus WP5 shows that the reality of implementation differs from this ideal scenario. The report highlights the shortcomings of the quantitative approach and the lack of policy discourse between the European and national levels as well as within the national level. What we see in several cases is a missing link from the level of policy implementation back to the top-down level. In particular, where monitoring and evaluation are not standard tools of policy making, there is a lack of discussion about the experience of implementing policies and the results achieved. A national discourse should also focus on discussing what has worked, why or why not, and what lessons can be learnt for future policies. This would also make it possible to fight the antigender movements on a solid basis.

The presentation will be structured as follows: Firstly, I will describe the differences between the approaches in the 2016-2020 and 2022-2025 ERA periods and the extent to which lessons learned have been taken into account in the design of the current instrument. Secondly, I will outline the role of monitoring in an ideal policy cycle and illustrate the aim and different purposes of monitoring in the context of gender equality in the ERA. Thirdly, I will discuss the relevance of monitoring for a policy discourse on gender equality, which provides the basis for linking policy implementation at European and national levels. This also provides the basis for actively confronting anti-gender movements and conservative currents on the basis of empirical evidence.

REFERENCES:

- Council of the European Union (2021a): Council recommendation on a Pact for Research and Innovation in Europe, 13701/21, Brussels, 19 November 2021, online available: https://era.gv.at/public/documents/4620/Pact_for_RI_26112021_EN.pdf [28.05.2024]
- Wroblewski, Angela (2021): Monitoring of ERA priority 4 implementation – update and final assessment, GENDERACTION D3.3, online available: https://genderaction.eu/wpcontent/uploads/2021/09/GENDERACTION_WP3_final_report.pdf [28.05.2024]
- Wroblewski, Angela (2023), First report on monitoring ERA action implementation at national level, GENDERACTION+ D5.1, online available: https://genderaction.eu/wpcontent/uploads/2023/09/GENDERACTIONplus_D5.1_First-report-on-monitoring-ERA-actionimplementation-at-national-level.pdf [28.05.2024]

Elizabeth Koier, Yvonne Lisman- van Leeuwen
University Medical Center Groningen, Netherlands

WHAT WE LEARNED FROM 6 YEARS OF MODERN RESEARCH ASSESSMENT

ABSTRACT

In the University Medical Center Groningen we have four different profiles for academic personnel since 2018, including a well trained promotion committee, narrative cv and DORA proof criteria. These profiles are based on the impact of someone's research: On other research, education, society & valorization or health care.

This policy is working well in many aspects:

1. There is more room for personalized decisions due to the narrative CV
2. The training of the promotion committee makes the promotion procedure well received
3. There is acceptance of the new DORA-proof research quality indicators

However, some of the fundamental goals are not met yet. Especially the goals on workload reduction, team formation and profile choices did not work out as planned.

We made an analysis of why this is the case based upon:

1. An employee survey
2. A follow up survey specifically on workload and social safety
3. Focus groups on both topics
4. Our experiences as academic career advisors, secretaries of the promotion committee and policy advisors

Our conclusion is that there are several fundamental reasons why policies and profiles created six years ago are not enough to reduce workload and equal recognition of the profiles in (the medical part of) academia.

1) Too much pressure on the individual in terms of their own research line, promotion and individual success. The result is:

- a lack of collaboration within departments (fear that it is not clear enough that the research line is someone's own)
- Pressure to grow to the top (full professor, associate professor with rights to decide if a PhD- student is ready for the defence)
- A CV must contain all tasks (research, education, societal impact, health care, outreach, leadership, funding success, supervision), but in the end the most important task for a career is success is research (and not in education, societal impact etc.)

2) A skewed composition of the academic workforce (more full professors than assistant professors)

- This results in fewer chances for young academics (professors have a much higher salary)

- An ageing workforce that is replaced by seniors without enough juniors to replace them
- Most (associate) professors do not really have a leading function with respect to their colleagues (they only have PhD-students and one or two postdocs, no assistant professors or others to lead)

3) Little recognition for leadership talents (brilliant academics are supposed to learn how to lead others, leadership talent is not seen as a key reason to promote a good academic to a leading position, although a severe lack of leadership talents may be a reason not to promote someone)

- No explicit time allowance for leadership tasks
- No clear leadership quality criteria

4) Because research evaluation is based on whole CV's, change in academic focus (to education, societal impact) is not appreciated because it is seen as a gap in someone's research CV.

- Keeping the work varied and challenging by changing academic focus for a period of time is therefore not recommendable

We want to gradually move towards the following situation:

1. Research assessments are reduced due to less focus on promotion. After some steps from postdoc to (senior) assistant professor, recognition takes place in principle within the same position. A position is kept interesting through horizontal development and also in all kinds of committees. We must get rid of the requirement that everyone must be full professor to be taken seriously.
2. Promotion criteria are replaced by expectations of someone in a position measured over 3-5 years (so one can also switch profiles without assessment of one's entire CV, making it possible to change focus without creating a gap in your cv).
3. Expectations and research assessment criteria on societal impact and education should also be about quality not only about ticking the box
4. There must be expectations and assessments at the team level, without forgetting the role of the individual
5. Time investment on different tasks must be made explicit and criteria must also be adjusted accordingly. There is a lower limit to the scalability of criteria. More attention should also be paid to participation in projects and therefore not being the initiator if one's focus is on another academic area. Now participating in a project hardly counts.

In the coming year the UMCG is working on these topics in varied committees. However there are some open questions for academics thinking about research assessment:

1. How to design assessments of the quality of education and societal impact
2. How to assess in time blocks instead of entire CVs
3. How to assess academic teams as a whole (with tasks in education, societal impact, research, health care) and the position of individuals within them
4. How to scale criteria to take other tasks into account and how recognize participation in projects in assessments
5. How to assess leadership and leadership talent and how to incorporate this into the promotion criteria

Kathleen Toepel
IfS Institut für Stadtforschung und Strukturpolitik GmbH, Germany

UNVEILING INNOVATION: USING INNOVATION BIOGRAPHIES IN EVALUATION PRACTICE - A REFLECTION

ABSTRACT

A wide range of RTI funding measures are implemented at regional level in European Structural and Investment Funds programmes. Funding is organised according to specific objectives in so-called priority axes. For the 2014-2020 funding period, Berlin's ERDF (European Regional Development Fund) programme included among other actions direct funding for R&D projects by companies conducted in collaboration with research institutes, support for innovative start-ups through venture capital funds and the establishment of application laboratories and validation centres at research institutions and universities. The specific objective of this programme is to intensify and expand the innovation activities of industry. The entire innovation process is to be strengthened. In addition to research and development, services, and the creative industries also play an important role in Berlin. A total of almost 600 million euros in eligible expenditure was available for this purpose, thereby making this axis the largest component of the programme. The actions under this priority axis covered all phases of the innovation process (applied industrial research, experimental development and production set-up / market launch) as well as finding co-operation partners through network funding (in the clusters of the Regional Innovation Strategy and within the cultural industries).

A multi-year, accompanying evaluation for the ERDF programme was conducted on behalf of the Berlin Managing Authority in the Senate Department for Economic Affairs, Energy and Public Enterprises. In accordance with Article 54 of Regulation (EU) No. 1303/2013 and Berlin's evaluation plan, the effectiveness of the funding, its efficiency and its impact of each individual priority axis were assessed. The accompanying evaluation of the 'Innovation' priority axis started in June 2016. An interim report was produced in 2018 (IfS 2018). The final report was published in 2023 (IfS 2022).

The study design for the evaluation of the priority axis was programme theory based (Funnel and Rogers 2011, Rogers 2014). The aim of the study was to reconstruct impact pathways and show how and under what circumstances the interventions work. In accordance with the evaluation plan, the study design was based on the understanding of the theory of change of ERDF funding as described in the 'Guidance Document on Monitoring and Evaluation', according to which external factors in addition to the intervention also have an impact on the results (European Commission 2014: 5). In this priority axis, the interplay between the individual measures and other external factors was examined during the course of the programme. The design thus corresponds to the complex objective of the priority axis ('strengthening the entire innovation process') and

the diverse individual measures.

As a novelty in evaluations, innovation biographies (adapted from Butzin et al. 2012) were chosen as the central method. The final report is largely based on the cross-evaluation of the innovation biographies. Innovation biographies are a new research approach that can be used to empirically capture knowledge dynamics in innovation processes from a spatial and sectoral perspective. Innovation biographies make it possible to model the process of knowledge generation to be modelled in concrete innovation processes, from the initial idea to the concrete form of a new product or service, production set-up and market launch. This approach takes into account changes in the theoretical and empirical debate on innovation (Rammert 2000, Crevoisier and Jeannerat 2009). At the same time, an exploratory approach enables the mapping of the influence of external factors in the same manner as that of the various support measures, thus facilitating a more comprehensive understanding of the mode of action.

The evaluation design, in the form of such innovation biographical case studies, allows for a thorough examination of the individual case. This is particularly evident when considering the specifics of the project, historical coincidences, and external influences. The development history of the projects in their interaction with the environment in which they are operating can thus be analysed.

A ‘panel’ of 23 innovation projects was utilized to map the respective innovation development. The cases were selected from different measures and funding years. The selection of cases was informed by an appropriate mix of sectors, fields of technology and company sizes, as well as types of organisation (enterprises, universities, non-university research institutions).

The accompanying evaluation of the ERDF allowed for a long-term study design over the entire funding period. Each selected project was interviewed once a year. The observation period for each individual case is at least four and up to six years in duration. In some cases, the history of each innovation is documented over a period of up to 10 years. This allowed a long-term perspective beyond the (limited) duration of the funded project. In addition to the annual interviews with the management or project leaders, exploratory interviews were conducted with key cooperation partners and investors.

An innovation biography was created for each case study, in which all significant aspects and factors influencing the development of innovation were presented and integrated into their context. This included an explanation of the impetus that led to the initial idea, the obstacles and difficulties encountered, a trajectory of knowledge development and the associated network of stakeholders. Furthermore, the chronological sequence of predecessor and successor projects was also considered. The innovation biographies were updated annually. Finally, the analysis of the innovation biographies was conducted in a joint manner, in accordance with the impact pathways of the theory of change.

The contribution presents experiences with this qualitative method, using vivid examples from the fields of green and digital transition. It clearly shows the interplay of different funding measures (including ERDF and other national and European funds) in the innovation process and the influence of political and regulatory frameworks and other external factors. The narrative style used in the innovation biography makes the impact mechanism of the funding visible, thereby facilitating the communication of evaluation results to stakeholders and a wider audience. The innovation biographies also provide an authentic account of the extended time periods required to realise the outcomes of the funding. In certain cases, this spans across several funding periods. The advantages and disadvantages of this approach and its applicability to the evaluation of transformation processes are discussed. In particular, this method can be used to capture the complexity of transformation processes at the micro level. Consequently, the paper contributes to the advancement of R&I policy evaluation frameworks and methods.

REFERENCES:

- IfS (2018): EFRE-Fˆrderung im Land Berlin 2014 bis 2020 - Evaluierung der Priorit ‰ tsachse 1 „Innovation". Zwischenbericht. Dezember. https://www.berlin.de/sen/wirtschaft/gruenden-undfoerdern/europaeische-strukturfonds/efre-foerderperiode-2014-2020/der-efre/operationelleprogramme/das-operationelle-programm-des-efre-2014-2020/zwischenbericht_evaluierung-pa-1.pdf
- IfS (2022): EFRE-Fˆrderung im Land Berlin 2014 bis 2020 - Evaluierung der Priorit ‰ tsachse 1 „Innovation". Endbericht. September. https://www.berlin.de/sen/wirtschaft/gruenden-undfoerdern/europaeische-strukturfonds/efre-foerderperiode-2014-2020/der-efre/operationelleprogramme/das-operationelle-programm-des-efre-2014-2020/endberichtevaluierung_pa_1.pdf
- Butzin, A. / Rehfeld, D. / Widmaier, B. (Hrsg.) (2012): Innovationsbiographien – R‰umliche und sektorale Dynamik.
- Crevoisier, O. / Jeannerat, H. (2009): Territorial Knowledge Dynamics. From the Proximity Paradigm to Multi-location Milieus. European Planning Studies. 17 (8), 2009, pp 1223-1241.
- European Commission (2014): Guidance Document on Monitoring and Evaluation.
- Funnell, Sue C., Rogers, Patricia J. (2011): Purposeful Program Theory: Effective Use of Theories of Change and Logic Models.
- Rammert, W. (2000): National Systems of Innovation, Idea Networks, and Comparative Innovation Biographies, TUTS-WP-5-2000. http://www.ts.tu-berlin.de/fileadmin/fg226/TUTS/TUTS_WP_5- 2000.pdf
- Rogers, P. (2014). Theory of change. Methodological briefs: impact evaluation, 2(16), 1-14

Anabela Santos, Francesco Molica, Andrea Conte
European Commission, Joint Research Centre, Spain
European Commission, Joint Research Centre, Belgium
European Commission, Joint Research Centre, Spain

DRIVERS OF ECONOMIC AND INNOVATION DIVIDE ACROSS EUROPEAN REGIONS: THE ROLE OF EUROPEAN FUNDS

ABSTRACT

The economic divide among EU regions poses significant challenges to achieve a cohesive and competitive European economy. Reducing regional disparities among EU regions can foster more inclusive economic growth, promoting equity and cohesion. Innovation plays a pivotal role in reducing economic gaps by fostering growth, driving productivity, and creating opportunities for inclusive development.

This study seeks to provide a comprehensive overview of the economic divide across EU regions over the last two decades, to explore its relationship with innovation and to understand the role of EU funds to close the gap. In the present study, the term economic divide refers to the distance or gap in economic conditions, measured by Gross Domestic Product (GDP) per capita, between different regions within the EU. To estimate the economic gap index we use the following data from Eurostat: (i) gross domestic product (GDP) at current market prices expressed in purchasing power standards [nama_10r_2gdp] and (ii) average annual population [nama_10r_3popgdp]. Values are transformed at constant prices (base 2015) using the Gross Value Added deflator available in ARDECO website.

The gap index ($\mathbf{GAP_{i,t}}$) is estimated using a similar approach than Aghion et al. (2005). As expressed in the equation below, where $\mathbf{Z_{i,t}}$ refers to regional productivity of region i in year t, measured by GDP per capita. $\mathbf{Z_t^{max}}$ refers to the maximum observed values of GDP per capita in the EU (between index) or within a country (within index).

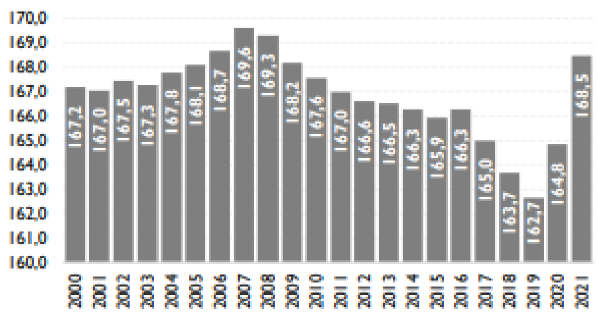
$$\mathbf{GAP_{i,t}} = \left[\left(\frac{\mathbf{Z_t^{max}} - \mathbf{Z_{i,t}}}{\mathbf{Z_t^{max}}}\right) \times \mathbf{100}\right] + \mathbf{100}$$

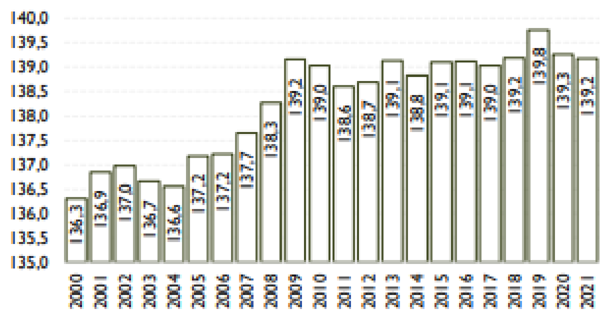
The gap index takes values between 100 and over 200. The gap index with a value of 100 corresponds to the region registering the maximum value in the EU or within a country in terms of productivity, meaning that this region is the benchmark to estimate the value of the gap. The higher the value of the index, the more distant the region is from the maximum value observed.

Figure 1 and Figure 2 report the average economic gap index over time, that is, the average regions' gap with the best EU and national best-performers, respectively. The economic gap at EU level has shown a decreasing trend since 2009 (Figure 1), except in the 2020 and 2021 during the Covid-19 pandemic. However, the convergence process prior to the Covid-19 appears to be far from steady or homogeneous.

FIGURE 1. ANNUAL AVERAGE ECONOMIC GAP (BETWEEN INDEX - DISTANCE TO EU MAXIMUM), 2000-2021

FIGURE 2. ANNUAL AVERAGE ECONOMIC GAP (WITHIN INDEX - DISTANCE TO MAXIMUM WITHIN COUNTRY), 2000-2021





Source: Own elaboration based on Eurostat data.

Our analysis is in line with previous work (see e.g. Viesti, 2021), which shows that a number of Member States have experienced a geographical concentration of economic activities over the past twenty years, leading to growing regional disparities. Another telling aspect is that the convergence process has followed an irregular pattern over time. At EU-level it increases until 2007 and then declines (until Covid-19 pandemic), while at national level it increases until 2009 and remains more or less stable in the following years.

There is an obvious relationship between innovation performance and economic divide at regional level. Our analysis illustrates this close link. Between 2000 and 2019, EU regions with the strongest innovation performance (so-called "leader" and "strong" innovators in the Regional Innovation Scoreboard – RIS – classification) have converged (in relation to the best EU performer) more than the other territories (Table 1). On the other hand, regions with a lower innovation capacity have reduced their economic gap less (emerging innovators) or even experienced a slight increase (moderate innovators).

TABLE 1. ANNUAL AVERAGE ECONOMIC GAP (EU INDEX - DISTANCE TO EU MAXIMUM) BY REGIONAL INNOVATION SCOREBOARD CLASSIFICATION (RIS 2023)

TABLE 2. ANNUAL AVERAGE ECONOMIC GAP (IN-COUNTRY INDEX - DISTANCE TO MAXIMUM WITHIN COUNTRY) BY REGIONAL INNOVATION SCOREBOARD CLASSIFICATION (RIS 2023)

Category	2000	2018	2019	2020	2021
Emerging	180.2	177.0	176.1	177.5	179.5
Moderate	167.3	166.6	165.4	168.1	171.3
Strong	162.2	156.5	155.6	157.7	162.1
Leader	153.0	147.4	146.3	149.0	154.7

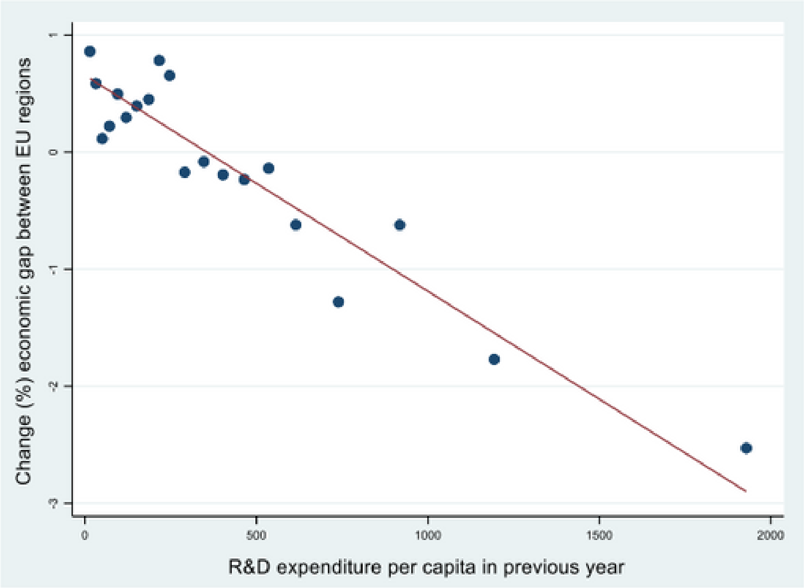
Category	2000	2018	2019	2020	2021
Emerging	180.2	177.0	176.1	177.5	179.5
Moderate	167.3	166.6	165.4	168.1	171.3
Strong	162.2	156.5	155.6	157.7	162.1
Leader	153.0	147.4	146.3	149.0	154.7

Source: Own elaboration based on Eurostat data and the RIS scoreboard

This trend is even more pronounced when seen through the lens of in-country disparities. Table 2 shows that regions with a weak innovation performance have diverged substantially from the richest region in their respective country over the period 2000-2019. Leader and strong innovative regions have instead improved their economic position vis-‡-vis the best performer in their country.

Regional disparities are also closely related to R&D expenditure. Overall, Figure 3 shows that the higher the R&D expenditure per capita in one year, the higher the reduction in regional disparities in the following year.

FIGURE 3. CORRELATION BETWEEN CHANGE IN ECONOMIC GAP BETWEEN TWO PERIODS (%) AND R&D EXPENDITURE PER CAPITA IN THE PREVIOUS YEAR (EUR), 2000-2021



Source: Own elaboration based on Eurostat data.

Note: The figure above is a binscatter constructed using panel data of 4,977 observations. Binned scatterplots provide an alternative way of visualizing the relationship between two variables (which does not imply causation), with a large number of observations, by computing the mean of the x-axis and y-axis variables within each bin and then creates a scatterplot of these data points.

After this contextual analysis, we use these measures of gaps as dependent variables in a three-stage estimation for systems of simultaneous equations, to assess how they influence each other, and how the different EU funds (Framework Programmes for Research and Technological Development, together with Cohesion policy) have contributed to close this gap. To carry out such an analysis, we use data from Eurostat, Cordis and the Open Cohesion Data Platform, to construct a panel data set covering 242 regions of the European Union over the period 2000-2021.

Paloma Martin, Doorenbos Pascalle
Research Executive Agency, Belgium

'PROJECT PATHWAY TO IMPACT' VERSUS 'KEY IMPACT PATHWAYS' IN HORIZON EUROPE CLUSTER 2

ABSTRACT

Horizon Europe, the 9th EU Framework Programme for R&I, has a budget of EUR 95.5 billion for the period 2021-2027. It succeeds Horizon 2020 which covered the previous MFF 2014- 2020. Given the economic and societal challenges that Europe is facing, it is, more than ever, essential to maximize the EU's budget effectiveness, to deliver tangible results and impacts. However, measuring the tangible impacts of research and innovation investments remains challenging.

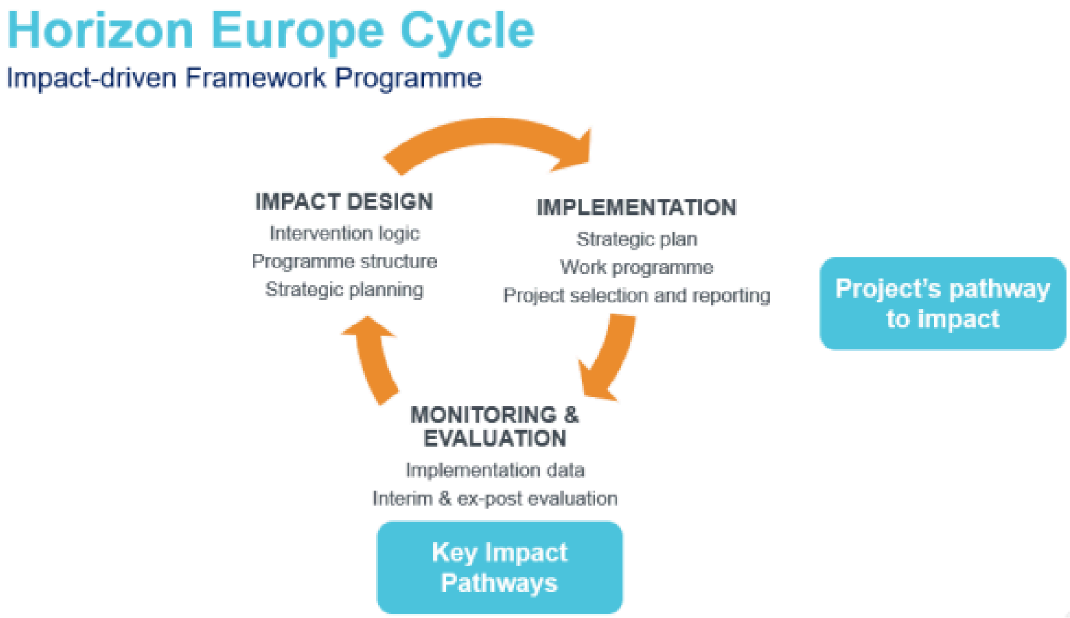
The interim evaluation of Horizon 2020 showed that the indicators system used for assessing the results and impact of Horizon2020 did not allow to tell the story of the diversity of impacts of the programme as a whole. As a consequence, a more systemic approach was needed.

The High-Level Group, chaired by Pascal Lamy, recommended in 2017 that the post-2020 EU R&I programme captures and communicates the impact better, through a comprehensive and centralised programme monitoring and evaluation system.

As a response, a new more impact-oriented framework, was set-up for monitoring and evaluating Horizon Europe. Which is enshrined in the Horizon Europe legislation (the so-called key impact pathways).

The Key Impact Pathways monitoring framework consists of a set of 27 indicators, structured around nine key story lines, to report at short, medium, and longer terms on the progress towards scientific, societal and economic impacts, in line with the specific and general objectives of the programme.

The Key Impact Pathways will be feed with implementation data coming from the implementation of the Programme and the Project's Pathway to impact under the following Horizon Europe Cycle:



Following the Horizon Europe Intervention Logic, the strategic orientations for R&I investments are defined in the multiannual Horizon Europe strategic plans which act as a compass for defining Horizon Europe's activities. The latter are operationalised through (annual/two years) work programmes which set out funding opportunities. The HE impact construct is thus largely using a holistic success measurement where the overall success of the programme is evaluated based on the success of its individual components.

Funded Horizon Europe projects are expected to bring benefits to society, the economy and science, first through their outputs and then, through their expected outcomes/results and impacts:

- Project outputs are concrete short-term outputs created during the implementation of Horizon Europe's projects, such as publications, innovative solutions, algorithms, new business models, prototypes, trained researchers, new infrastructures, new standards etc.
- Expected results (outcomes) are mid-term effects of Horizon Europe projects such as uptake, diffusion, use and deployment of the project's results by direct target groups. They are directly linked to the actions supported and allow observing direct outcomes during implementation and not only at the end of a project. Thus, they provide a strong basis for later evaluation activities.

Expected impacts are longer-term effects on society, the economy and science enabled by the outcomes of the R&I investments. This paper assesses the Project Impact Pathways on the basis of an internal analysis that run since March 2023. The study looked at all the signed Horizon Europe projects of HE CLUSTER 2.

The specific objectives examined how the HE projects (SSH) have conceptualised their impact pathways. In particular:

- To assess whether a logical ‘impact’ pattern is attained at project level;
- To collect information about the types of impacts proposed (scientific, economic, societal);
- To examine the outcomes at topic level and the effect in relation to the impact pathway at project level;
- To analyse the relationship between the impact pathway against related topic, as well as the measurement construct to examine the indicators proposed and the type of indicators proposed.

The underlying idea of the analytical work was first and foremost to support CL2 applicants, notably via the interest of the NCPs in the SSH topics and to support the work of the Executive Agency Project officers. The results of the exercise might also be useful for Policy work and the larger monitoring and evaluation community.

A qualitative assessment run from March 2023 (cut off date) on a 136 EU financed projects. The Impact section was analysed, and the information retrieved on a spreadsheet. This allowed for the collection of the structured data that underwent different sets of statistical correlations and frequency analysis. To quality control the data, the information collected on the spreadsheet was digitally triangulated. The findings show that most of the topics are ‘closed’ topics i.e., with very defined top-down outcomes which the projects need to accomplish in full. All projects have indicators, and they are mostly QUALITATIVE indicators. Most of the projects DO NOT provide baselines OR Means of Verification.

There is also a low positive correlation between the way outcomes are drafted in the individual workplan topics, and the ´ outcome indicator: quantitative/qualitative ª (as a proxy of project impact pathway quality) across all three CLUSTER 2 destinations. In the 3 destinations, the frequency of ‘yes’ answers in ‘baselines’ and ‘means of verification’ is higher when the ‘outcome indicator is ‘yes, quantitative/qualitative’ compared to when is only ‘yes, qualitative’. This could mean that for the quality of the project impact pathway to increase (measured with proxy outcome indicators (quantitative/qualitative), baselines and means of verification increasing) special attention needs to be paid to the outcome typology and level of prescription. To conclude, providing a varied and ‘open’ outcome base, might have a direct effect on the quality of the project pathway to impact. Disclaimer: All opinions expressed in this paper are so current author opinions and do not reflect the opinions of the European Commission or the Research Executive Agency. The author opinion is based upon information considered reliable by the author, but either the European Commission nor the Research Executive Agency are responsible for its content and do not warrant its complete accuracy.

Thomas Palfinger
LBG OIS Center, Austria

TURNING MISSION INTO ACTION: THE CANCER MISSION LAB

ABSTRACT

This contribution gives insights into the conceptualization, planning and results of the Austrian Cancer Mission Lab (CML) carried out by the Open Innovation in Science Center of the Ludwig Boltzmann Society (LBG OIS Center). The CML is a pilot instrument funded by the Fonds Zukunft ÷sterreich (FZ÷) to implement mission-oriented funding, aligning with the European Cancer Mission within the Austrian context. The CML is therefore one of the national efforts to anchor the EU Missions in the national context and to implement a mission-oriented funding logic. This contribution will present the considerations that were central to the planning and development of the CML and ultimately shaped its concrete design. Against this background we had to solve two main challenges 1) operationalizing the mission orientation against the background of the funding instrument and 2) enabling transdisciplinary collaboration through the selection procedure.

The CML has a funding volume of 1.5 million to fund 2-4 projects for max. 36 moth. It is necessary for different stakeholders from the science, civil society, the healthcare sector to collaborate on a challenge they have identified during the process. For this reason, different stakeholder groups were eligible to apply, which had to be considered in the decision making for the funding instrument (i.e. selection process, language, length etc.). The persons and ideas were selected by a six-member committee, two of whom are assigned to each of three forms of expertise: scientific expertise, system-specific expertise (Austrian healthcare system), lived expertise.

To establish a mission-oriented funding instrument, we had to operationalize the meaning of mission to an extent that we could use it to make decisions. Here we were guided by a study of Wagner 2021, where development paths for mission-oriented research funding were provided. This resulted in a simplified idea of mission orientation, which made it easier for us to operationalize it within our framework. Mission orientation was reduced to three central components:

- Intentionality: The mission works towards a clearly defined goal. A successful mission solves a problem or helps meet needs. For communication and engagement, this means that a clear shared vision of the mission's goal must be formulated and conveyed so that everyone can act accordingly.
- Purposefulness: The mission is a cooperative endeavor with clearly defined tasks and a clear timeframe. The goals are ambitious and therefore inherently risky, but the stakeholders are convinced of their benefits. For communication and involvement, this means helping new stakeholders focus on the common goal and involving them accordingly.

- Reflectiveness: Due to its cooperative nature, the mission involves a variety of actors and instruments. For communication and involvement, this means an increase in complexity, many different activities need to be orchestrated, and groups need to be involved. The progress of the mission must be transparently monitored and communicated.

This understanding of the mission consisting of intentionality, purposefulness and reflectiveness accompanied us in the design of the CML as well as the applicants in the way they had to navigate through the different phases and documents of the call. As mentioned, potential applicants were from diverse institutions and backgrounds, which resulted in the second challenge of the instrument: how do we design a fitting funding instrument for very diverse actors with differing goals?

As a result, we structured a three-phase process designed to enable easy interactions with the funding instrument, while still attract relevant actors and ideas. Another very important aspect of the process was to generate functioning transdisciplinary teams at its end:

1. Phase 1: Searching Persons (March 15 - May 15, 2024) - Interested participants submit a motivation letter outlining their vision, collaboration approach, communication strategy, and relevant expertise. Participation in a mandatory consultation session is required to ensure alignment with the program's objectives.
2. Phase 2: Identifying Ideas (July 10-12, 2024) - Selected candidates participate in a so called Ideas Lab, forming interdisciplinary teams to develop preliminary project sketches. The lab fosters a creative environment with the support of mentors and moderators, culminating in the selection of 2-4 project ideas for further development.
3. Phase 3: Turn Ideas into Projects (July - November 2024) - Chosen teams refine their projects with expert guidance, creating detailed proposals including a "Theory of Change," budget, and timeline. A kick-off workshop and continuous support ensure robust project planning, leading to the final submission by November 8, 2024.

Results and learnings of the Process will be presented at the fteval conference in December 2024. As of May 18, over 120 people have taken part in a consultation for the CML, 78 of whom have submitted a motivation letter for the first phase.

REFERENCES:

- Wagner, Isabella (2021) Programmmanagement und Kommunikation in der missions-orientierten Forschungsf˚rderung. Am Beispiel "Stadt der Zukunft". Technischer Bericht. Bundesministerium f¸r Klimaschutz, Umwelt, Energie, Mobilit ‰ t, Innovation und Technologie (BMK). Wien.

Stephanie Daimer, Matthias Weber
Fraunhofer Institute for Systems and Innovation Research ISI, Germany
Austrian Institute of Technology AIT, Austria

HOW THE RECENT STI POLICY SHIFT TAKES SHAPE IN THE EU RESEARCH FRAMEWORK PROGRAMME

ABSTRACT

The "Ex-post Evaluation Study of the Relevance and Internal Coherence of Horizon 2020 and its Policy Mix" (European Commission 2023) was the only study among the set of 12 Horizon 2020 expost evaluation assignments asked to study Horizon 2020 strategic processes, including programming, internal and external consultation and co-creation, foresight and policy feedback. The study had the task to not only evaluate the processes during the runtime of the programme between 2014 and 2020, but the relevance and internal coherence of all elements of the policy mix (understood in a broad sense as e.g. brought forward by Rogge and Reichhardt (2016)).

The evaluation study was challenged by the fact that the "shift" of science, technology and innovation (STI) policies towards challenge-orientation and transformative change increasingly was adopted by Horizon 2020, resulting in the orientation towards broader policy objectives and higher ambitions, a reconsideration of intervention logics and a change of activities and instruments. Horizon 2020's orientation on societal challenges had also led to changes in the programming processes with a partly amended process for consulting stakeholders and involving advisory boards (via so called scoping papers) as well as a more integrated approach of different DGs, and the invention of multi-annual programmes. This new multi-annual approach was later coined as the strategic programming process, which came on top of the annual work programme updates, which continued to exist. This process innovation is just one example among others in Horizon 2020.

Against this backdrop, with this contribution, we want to share insights and experiences from the evaluation of relevance and internal coherence of Horizon 2020 regarding three questions:

i) How did Horizon 2020 strategic processes of policy development and learning support the internal coherence of the programme?

ii) How relevant and suitable was the portfolio of instruments and activities to the new and more ambitious goals pursued?

iii) To which extent is this shift in ambition not only reflected in the programme, but also in the evaluation requirements and practices?

As we will show, the findings from this analysis are of relevance also for Horizon Europe

and for FP 10, as some challenges continue to exist. Further, we noticed that the approach to evaluation at least for the current interim evaluation of Horizon Europe has changed in a way that the evaluation of relevance and internal coherence is limited to intervention logics, instruments and activities, leaving strategic processes aside. So, with this analysis we seek to underline the value of a process perspective in framework programme evaluation in particular against the backdrop of the requirement to evaluate the transformative potential of the programme. One of our findings in this regard points to the limitations of a summative evaluation approach when analysing processes, as the organisational learning remains limited as compared to the opportunities for learning in a formative evaluation setting.

Scope of the evaluation: The evaluation study covered the whole Horizon 2020 programme (2014-2020). It was performed in line with the Better Regulation guidelines as applicable to all EU expenditure programmes. The focus of the evaluation was on the relevance and internal coherence evaluation criteria. The 'coherence' criterion splits broadly into internal and external coherence, and it is internal coherence that was considered in this study. Evaluating 'effectiveness', 'efficiency' or 'EU added value' was not part of this assignment and was addressed in related Horizon 2020 impact area studies. In the context of this specific paper, and building on the corresponding baseline definitions in the Better Regulation toolbox, 'relevance' and 'internal coherence' of the policy mix and the corresponding processes established are defined as follows:

- Internal coherence of processes: A set of process characteristics, expected to support the successful design of the programme.
- Relevance of the policy mix (see the next section for a definition): The degree to which the policy mix addresses the policy objectives; the adequacy in addressing the needs of target groups.

In the evaluation study on the basis of which we have developed this paper, other aspects of relevance and internal coherence of the policy mix are addressed, so these two represent a selection for this paper.

We will build our analysis of the first two research questions of this paper on the empirical data collected for the evaluation study in a multi-method approach during the period August 2021 to September 2022. The mix of data collection methods comprises qualitative and quantitative text analysis, programme data analysis, a broad case study approach policy workshops and stakeholder consultation. These data were analysed using a triangulation approach. A broad understanding of policy mixes: The most basic definition of a policy mix is ‘a combination of several policy instruments. Several scholars have argued that this is better called ‘an instrument mix’, while the term ‘policy mix’ contains more complex processes, in which these instruments emerge and interact (Flanagan et al. 2011). Furthermore, policy-mixes are characterised by overarching long-term strategies which define their objectives (Rogge and Reichardt 2016). It is also stressed, that policy mixes often evolve over time, with instruments being amended or ended, and new instruments added; all this affects the interaction of the instruments in the mix – and very often the outputs, outcomes and impacts.

In our study, we follow the broad conceptualisation of policy mixes (Rogge and Reichhardt 2016) in terms of goals, instruments and processes as follows:

- The strategic orientation and overarching rationales when the programme was conceived, and how they evolve during the runtime of the programme;
- The instruments and activities, as well as their internal coherence and complementarity; • The strategic processes by means of which goals are defined, content orientations of programmes in the portfolio ensured, and reciprocal (coherent) adaptation and learning enabled (as opposed to operational processes ensuring efficiency and effectiveness of the implementation, which are not part of our study).

The "shift" in STI policies and approaches to the evaluation of transformative policies The "shift" in STI policies towards addressing societal challenges has been discussed as missionoriented innovation policy (cf. e.g. Mazzucato 2017, Larrue 2021), transformative innovation policy (cf. Diercks et al. 2019, Schot and Steinmueller 2018), or transformative mission-oriented innovation policy (cf. Edler et al. 2024, forthcoming). It started to manifest in EU R&I policy with the Horizon 2020 framework programme and its programmatic re-orientation towards addressing societal challenges, although it remains to be debated how and to what extent this re-orientation actually was realised in implementing Horizon 2020.

For the purpose of our analysis of the third research question, we will use a set of characteristics describing the essence of the new generation of STI policies, which is tailored to the analysis of a programme. It will include, e.g., directionality, a broad understanding of innovation, inclusion of a broader range of innovation actors, and the addition of the demand-side activities to supply-side instruments. Using these characteristics, we will analyse how the components of the policy mix (as outlined above) have changed over time in Horizon 2020.

As a second component to assess the third research question, we will need to draw on existing approaches to the evaluation of transformative innovation policies (e.g. Janssen et al 2022, Haddad and Bergek 2023), and derive a set of criteria to assess how the evaluation approach of the European Commission has changed over time.

We find that Horizon 2020 can be delineated into three stages: 1. A first programme phase without explicit intervention logic, 2. A second phase with a rather traditional innovation and impact-oriented logic (following the interim evaluation and the high-level Lamy report), 3. A third phase, mainly in response to the COVID-19 research initiatives and the Green Deal Call, attempting a socially transformative logic. These logics are underpinned by stronger strategic orientation and guidance, reflected not only at the political but also at the operational level. At the same time, the policy mix was partially adapted, however not with the rigour needed.

The evaluation study had a particular focus on programming processes and functional processes supporting programming such as foresight, policy learning or policy feedback. These processes were developed further during the period of the programme, and in particular the strategic programming took shape, however was found to still lack efficient mechanisms to re-orient and re-prioritize. Functional processes gained in relevance but were until the end of Horizon 2020 not deeply institutionalized. In other words, process changes were lagging behind changes in intervention logic and instruments.

The reasons for this delay are manifold, but the need for increased collaboration between DGs and directorates seems to foster the development of a comprehensive and rather complex programming processes, while the self-interests and logics of action of DGs and directorates, the lack of institutional memory, or the different degrees of influence of the various stakeholders were found to inhibit process changes supporting a major policy shift.

In the meantime, i.e. in Horizon Europe, some of the early learnings from Horizon 2020 have been taken up and institutionalised.

Although our H2020 evaluation is still traditional in principle, we come to the conclusion that it points to the need for new forms of embedding evaluations into processes of policy learning involving a much wider range of Commission actors as well as external experts and stakeholders than before. This results from the need for increased collaboration between DGs and directorates as well as external experts and stakeholders to foster the development of comprehensive and rather complex programming processes. It has become clear by now that the ambitious policy goals formulated cannot be achieved without complementary actions not only of other Commission services, but also of national and regional policies, and the actions of private and third sector agents. EU R&I policy, and the framework programmes in particular, contribute to the unfolding of transformations, and the understanding of their roles should be framed accordingly. Although the relevance of evaluation for policy learning for subsequent framework programmes has grown immensely over the past years, a formative approach and hence change in evaluation culture might further support this process in the future.

REFERENCES:

- Diercks, G.; Larsen, H.; Steward, F. (2019). Transformative innovation policy: addressing variety in an emerging policy paradigm. In Res. Policy, 48, pp. 880-894.
- Edler, J., Matt, M., Polt, W., Weber, M. (2024, forthcoming): Transformative Mission-Oriented Innovation Policies, Edward Elgar, Cheltenham.
- European Commission (2023). Evaluation study on the relevance and internal coherence of Horizon 2020 and its policy mix – Final Report. Brussels.
- Flanagan, K.; Uyarra, E.; Laranja, M. (2011). Reconceptualising the 'policy mix' for innovation. In: Research Policy 40 (5), pp. 702-713.
- Haddad, Carolina, R.; Bergek, Anna (2023): Towards an integrated framework for evaluating transformative innovation policy. In Research Policy Volume 52, Issue 2.
- Janssen, Mathijs J.; Bergek, Anna; Wesseling, Joeri H. (2022. Evaluating systemic innovation and transition programmes: towards a culture of learning. PLOS Sustain.Transform., 1, Article e0000008
- Larrue, Philippe (2021): The design and implementation of mission-oriented innovation poli-cies. A new systemic policy approach to address societal challenges. OECD. Paris (OECD Sci-ence, Technology and Industry Policy Papers, 100). Available online at https://doi.org/10.1787/3f6c76a4-en.
- Mazzucato, Mariana (2017): Mission-Oriented Innovation Policy. Challenges and opportuni-ties. RSA Action and Research Center. London. Available online at https://www.thersa.org/globalassets/pdfs/reports/mission-oriented-policy-innovation-report.pdf, checked on 12/6/2019.
- Rogge, Karoline S.; Reichardt, Kristin (2016). Policy mixes for sustainability transitions: an extended concept and framework for analysis. In Res. Policy, 45, pp. 1620-1635.
- Schot, Johan and Steinmueller, W. Edward (2018), 'Three frames for innovation policy: R&D, systems of innovation and transformative change', Research Policy, 47 (9), 1554-67.

Michael Dinges, Gaîtan Coatanroch, Christiane Kerlen, Kathleen Toepel, Vitaliy Soloviy
AIT Austrian Institute of Technology , Austria
Technopolis Group , France
Kerlen Evaluation, United Kingdom
Kerlen Evaluation, United Kingdom
AIT Austrian Institute of Technology , Austria

HORIZON EUROPE AND THE GREEN TRANSITION: EVALUATION EVIDENCE AND POLICY CONCLUSIONS

ABSTRACT

The European Green Deal (EGD) is Europe's adapted growth strategy aiming to transform the EU into a fair and prosperous society, with a modern, resource-efficient, and competitive economy, where there are no net emissions of greenhouse gases (GHG) in 2050 and where economic growth is decoupled from resource use. In addition, the Green Deal emphasises the need to protect, conserve and enhance the EU's natural capital, and protect the health and well-being of citizens from environment-related risks and impacts. At the same time, the EGD aims to make this transition just and inclusive, by putting people first and paying particular attention to regions, industries and workers that will face the greatest challenges. As a new EU policy initiative, launched in 2019, the EGD sketches out all elements for the conceptualisation of a Green Transition in Europe.

Against this policy background, which puts the Green Transition at the center stage of Europe's growth strategy, this paper analyses by which means and to which extent Horizon Europe and its key instruments (Work Programmes, Partnerships, EU Missions) and governance mechanisms support the ongoing economic, societal, and industrial transformations towards a twin Green and Digital Transition.

The paper draws upon the main findings of a back-to-back evaluation of Horizon 2020 and Horizon Europe Framework Programmes. The analysis focuses on the key functions that the FP is able to take over as well as their limitations.

Based upon the main findings of the study, the paper puts forward learnings and perspectives on the future development of the European Framework programmes, and R&I policy making perspectives with a transition ambition.

Methodological approach: The evaluation study commenced its work by elaborating working definitions for a Green Transition and a methodological framework to consider the specific challenges of a Green Transition in relation to the instruments and actions set out in Horizon 2020 and Horizon Europe. The evaluation followed the main principles of theory-based evaluation (Chen 1990; Weiss, 1997; Rogers, 2007; Funnell & Rogers, 2011), and developed theories of change that related 1) the general and specific needs/challenges of

the Green Transition, with 2) the interventions of the related parts of the Framework Programmes and the European Partnerships. To answer the evaluation questions of the study in relation to the interventions in scope, the selected methodological approach mixed various data collection and data analysis tools. The different tools mobilised throughout the evaluation allowed to collect evidence to answer the various evaluation questions considered under this evaluation.

To analyse to which extent Horizon Europe has induced processes for a Green Transition, the evaluation uses the concept of the Multi-Level Perspective (MLP) and the embedded concept of transformative outcomes, which has been considered as guidance for the analysis of the interventions and operationalised in the survey design and the case studies.

To disseminate the findings of the study and validate the conclusions of the study, two policy workshops with representatives from the European Commission and Partnerships were conducted based on evidence collected.

Main Findings: The European Green Deal (EGD) has significantly shaped the design of Horizon Europe's work programmes, particularly in Clusters 4, 5, and 6, which address digital industry, climate, energy, mobility, food, bioeconomy, natural resources, agriculture, and environment. Horizon Europe allocates 35% of its budget to climate objectives, with Clusters 5 and 6 playing pivotal roles, aligning closely with Green Transition goals as approximately 90% of their projects support related themes. The programme's increased outcome orientation is evident, focusing on long-term impacts and strategic goals. Enhanced strategic coordination between different Directorate-Generals (DGs) within the programme aims to improve coherence and maximise the programme's impact.

Regarding priority setting and outcome orientation, Horizon Europe's design and prioritization processes, influenced by the EGD, emphasize long-term impacts and strategic goals. The programme has introduced mechanisms like the Key Impact Pathways (KIPs) to reorient from an activity-driven to an impact-driven framework, ensuring that projects align with broader policy objectives and longterm societal needs. This approach is one of the programme's strengths, allowing for a comprehensive coverage of emerging themes related to the Green Transition. However, the scale of Horizon Europe's contribution to the Green Transition needs clearer definition. While using the Do No Significant Harm principle, and integrating Cross-Cutting Specific Issues are good strategies, they fall short in addressing the necessary scale needed for fostering a Green Transition. In addition, the evaluation witnessed that the programme's impact focus also limits early-stage knowledge generation, which is important for Europe's future transformative potential.

The governance mechanisms of Horizon Europe have been enhanced to improve strategic coordination between DGs. This includes a co-creation approach in work programme design, involving a Steering Board composed of Director-Generals and an Executive Committee of Directors, which fosters better integration and alignment of R&I activities across different thematic areas. This structure should help mitigating fragmentation across programme parts. However, there are challenges in portfolio management between different parts of the Work Programme, requiring more effective thematic coordination.

Furthermore, creating synergies between Horizon Europe and other EU funding programmes remains a challenge. While the programme aims to strategically link its activities with other funding mechanisms to maximize impact, the study results show that proactive coordination and synergy creation mechanisms are limited, posing a significant barrier to the programme's effectiveness.

Horizon Europe's contribution to the Green Transition is marked by a strong focus on technological and industrial transitions. From a Multi-Level Perspective (MLP), the study analysed the extent to which Horizon Europe contributed to the Green Transition through three key macro processes: building and nurturing niches, expanding and mainstreaming niches, and opening up and unlocking regimes. The survey results indicate that Horizon Europe supports to a large extent the emergence of new fields of innovation by funding experimental projects and technological advancements. The programme also contributes to mainstreaming niches but actual impact on institutionalizing new strategies and norms is inconsistent. Horizon Europe’s contribution to challenging entrenched regimes is limited. While supporting projects that challenge existing norms, only a limited number of survey respondents expect high impact. This underscores the difficulty in disrupting established systems and the need for more research into effective transitioning processes.

Projects in Clusters 5 and 6 have shown a predominant focus on advancing specific technologies, but broader issues of societal change, stakeholder alignment, and behavioral shifts have been missing to some extent. This focus on technological advancements, while a strength in driving innovation, also serves as a limitation by placing less emphasis on the necessary socio-cultural and systemic transformations. The programme still faces notable challenges in societal integration and stakeholder engagement, as inclusivity gaps persist. International cooperation is at a lower level than in Horizon 2020, limiting broader collaboration despite a strategic focus on climate challenges. The programme's boundaries are defined by its focus on technological and industrial transitions, which, although critical, need to be complemented by efforts addressing socio-cultural changes and systemic reconfigurations.

Conclusions: The evaluation study has provided a wealth of insights into its impact and role in facilitating the Green Transition. The paper will elaborate on several conclusions that have emerged. Among these are:

An encompassing definition of the Green Transition: A clear understanding and definition of the Green Transition are crucial for increasing the directionality of the programme. The programme should expand its focus beyond climate change to include biodiversity protection, resource conservation, and pollution reduction. A more holistic approach could ensure that all environmental aspects are considered and addressed.

Addressing structural barriers and complex societal challenges: The programme's focus on technological and industrial advancements often overlooks essential socio-cultural changes and systemic reconfigurations necessary for a holistic transition. This technological bias limits its ability to address deeper societal transformations. For fostering the Green Transition, the programme must tackle deep-rooted structural barriers and complex societal challenges to facilitate systemic changes in the economy and society. Mechanisms to integrate varied objectives and stakeholder needs are essential. Acknowledging competing interests, power asymmetries, and diverse knowledge bases can lead to more inclusive and effective Green Transition strategies, when taking into account in the programme design. Engaging stakeholders through a multiactor approach and ensuring their needs are explicitly addressed in call designs are crucial steps. As the Green Transition depends on interventions beyond technological and market solutions, this includes considering more thoroughly diverse interventions beyond technological advancements, including a stronger focus on behavioural change of actors, social innovation, new governance models, and business practices. An interdisciplinary and transdisciplinary integration can lead to more comprehensive and sustainable solutions.

Balancing Phasing-Out, Upscaling, and Down-Scaling Activities: Greater consideration should be given to achieving a balanced approach in the Green Transition by properly mixing the phasing out of obsolete activities, upscaling necessary industries, and downscaling those incompatible with desired sustainable futures. Much of the research and innovation activities have been additive to the existing state-of-play, rather than transformative. It is crucial to question existing paradigms, and also conduct research into alternative futures, basic values, and beliefs to ensure a holistic transition towards sustainability. The Framework Programme could place greater emphasis on projects that aim to disrupt entrenched regimes and promote systemic change. This involves supporting initiatives that challenge existing norms and practices, elaborate new regulatory frameworks, and encourage innovative business models. Providing funding and resources for pilot projects and experimental approaches can help identify effective pathways for systemic transformation.

Improved Thematic Coordination and Portfolio Management: Thematic coordination across different parts of Horizon Europe should be enhanced to reduce fragmentation and improve coherence. Implementing a robust portfolio management system that tracks and integrates projects across various clusters and partnerships will help identify synergies and prevent duplication. Regular strategic reviews and adjustments based on portfolio performance and emerging needs can ensure a more dynamic and responsive programme.

Overcoming fragmentation and creating effective synergies: Improved responsiveness between the research and innovation domain and sectoral policymaking is vital. This includes focusing on structural barriers, societal readiness, market readiness, and new regulations and standards, to ensure innovations can effectively transition to market applications. Addressing complex challenges, governance fragmentation, policy failures, and future constraints requires systematic efforts. Active coordination between Horizon Europe and other EU funding programmes is necessary to maximize synergies and avoid duplication. Strategically linking activities with other funding mechanisms can enhance the overall impact and coherence of the programme. This can be achieved through better alignment of eligibility criteria, evaluation processes, and timelines across different funding instruments. Establishing dedicated coordination units or platforms to facilitate cross-programme collaboration and information sharing can help maximize the impact of combined funding efforts.

Matthias Weber
AIT Austrian Institute of Technology, Center for Innovation Systems & Policy UniversitÈ Gustave Eiffel, UMR-LISIS, Austria/France

STRATEGIC APPROACHES TO FUTURE-PROOF R&I POLICY

ABSTRACT

EU policy has been and will continue to be confronted with a wide range of potentially disruptive developments. They include emerging technologies likely to entail major consequences for society and economy (e.g. artificial intelligence, human enhancement, synthetic biology) to challenges arising from a new global order or from societal fragmentation and polarisation. These developments widen the range of future possibilities and scenarios, deepen the dependency of Europe's future on developments beyond its political influence, and call for a reframing of Europe's strategic options, not least in relation to R&I. This raises the question of how to make use of the forward-looking knowledge we have, and how to cope with the limitations of that knowledge when defining future visions, strategic priorities and policy mixes for European R&I policy.

Against this background, the objective of the paper is i) to systematically characterise the problem-solution space (Wanzenbˆck et al. 2020) of possible future disruptions and ii) to propose guiding models and strategic options for R&I policy (and beyond) to address different types of future problem-solution configurations.

The insights presented in this paper were developed as part of a strategic foresight project aiming to explore orientations for the 2nd Strategic Plan of Horizon Europe and the next European framework programme for research and innovation (Weber et al. 2023). This project drew on a combination of scenario development in expert group workshops, involving experts from science, industry, government and civil society, and extensive online surveys and consultations with wider stakeholder communities.

Characterising the problem-solution space of future disruptions: During the past ten to fifteen years, Europe and the world have been confronted with important social, economic, technological and global developments of a disruptive nature. While there is consensus that we need to strengthen preparedness for such disruptive effects, we are still attached to a strong belief in the ability to shape the future of Europe and its position in the world. We argue that there a need to re-balance the relationship between a defensive policy approach, aiming to strengthen preparedness for and adaptiveness to the unexpected, and an offensive and proactive policy approach, driven by the belief in the pursuit and achievement of (transformative) goals and missions. Next to these two extremes, we argue, a more catalytic role of R&I policy in shaping future transformation pathways should be pursued.

This, however, requires first of all to better understand the nature of potentially disruptive

developments and of the conditions for pursing a more or less targeted policy approach. We propose three critical dimensions to characterise this problem-solution space:

- The level of complexity and uncertainty of disruptions, i.e., the nature of the problemsolution space associated with potential disruptions.
- The level of power and autonomy to shape the future, i.e., Europe's ability to be the 'master of its own destiny', i.e. whether Europe is in a position to shape the future of these disruptive areas according to its normative ambitions and will.
- The level of consensus over the desired future(s) to be pursued, which reflects the challenge of first understanding and systematising the problem-solution space and second devising processes of societal decision-making, whether through market competition or through political and administrative mechanisms.

In view of these challenges, there are different basic strategies that the EU could pursue. First, the EU could take strongly coordinated and targeted actions of the kind that governments take in states of emergency ("war path"). For such a strategy to be meaningful and ultimately successful there need to be high levels of consensus and control over resources and options, and low levels of uncertainty about the problem-solution space and the desirable path to be pursued.

These conditions, however, are rarely met in practice. Successful implementation of a path would be easier if the EU had the power of autonomously shaping its future. Realistically, and recognizing the levels of interdependence of the EU in the global economy, we need to consider other paths, for instance, focusing on "selective cooperation" and what we shall call a "crowd path". In the first case, the EU and a limited number of partners are able to shape the global context to improve possibilities of turning the disruptive challenges into desirable outcomes. In the second "crowd" path, the whole (or most) of the international community aligns with the EU goals with regard to a disruptive challenge to jointly address it in a harmonised manner. It depends on the characteristics of the problem-solution spaces which of the three paths is most suitable from a European perspective.

The situation is further complicated, when there is no consensus over the problem-solution space, and targeted, directional interventions are contested. In the absence of consensus over the problem-solution space, issues of fairness and balance across alternative agendas are the key concerns for R&I policy makers, often taking precedence over directionality, coordination and policy coherence. In the paper, different possible configurations of the three distinctive dimensions of the problem-solution space will be illustrated by recent foresight work on disruptive developments explored in EU foresight projects in support of the implementation of Horizon Europe and the preparation of FP 10.

Roles of R&I policy in the context of future disruptions: three base approaches What do these considerations imply for European R&I policy on matters of future disruptions? To start with there is not a one-size-fits-all approach to addressing disruptive areas, but the roles of EU policy in general, and of EU R&I policy in particular, need to take into account the types and the nature of paths that seem most suitable for each specific disruption. In what follows we will reflect upon alternative R&I policy roles that might be pursued to complement current ones, in order to better cope with disruptive developments.

Since its inception in the 1980s, much of the EU's R&I policy, and in particular its framework programmes for research and innovation, has concentrated on an enabling role for science and technology development, in order to contribute to the strengthening of the competitiveness of European firms and, more lately, help address societal challenges. This enabling role of the EU's R&I policy is also highlighted in the European treaties, and it continues to be very important.

In recent years, and more specifically with the definition of major societal challenges as one of the three pillars of Horizon 2020, the EC has indeed emphasised what could be termed a transformative role of R&I policy. In this, R&I agendas are defined as pursuits of solutions in response to societal challenges, but run the risk of over-stretching the means of R&I policy.

Under such conditions, a more modest take on R&I policy seems appropriate, where instead of a transformative role one could speak of a catalytic role. For the latter, the key task of R&I policy consists of testing viable configurations of technological, social, behavioural, organisational and institutional changes, before they can be scaled and taken up more widely through a sound orchestration with other sectoral and cross-cutting policies as well as the strategies of industrial and societal stakeholders at different levels, from European to local.

As with the types of configurations of the problem-solution space of future disruptions, the different policy strategies and roles to respond to them – enabling, transformative, catalytic – will be illustrated by examples from current EU R&I policy for the second phase of Horizon Europe and the launch of the next framework programme FP 10. For instance, EU missions run the risk of relying too much on a transformative role of R&I policy, while it strongly depends on sectoral strategies, thus suggesting a rather catalytic role for R&I policy. Similarly, the influence of disruptive technologies in artificial intelligence or human enhancement, it seems, are driven largely by developments outside of Europe, suggesting a combination of a selective cooperation strategy that combines early regulatory efforts with a rather exploratory and enabling R&I policy.

Some normative policy orientations: Beyond leadership in the traditional sense

The different examples of potential future disruptions and their positioning in the problem-solution space in conjunction with the possible roles that EU R&I policy can assume point to a need to reconsider the strategic ambitions of EU R&I policy. These observations point to possibilities and limitations of striving for what is commonly called 'European leadership', but – more fundamentally – also to the need to reconsider the strategic ambitions that might be pursued beyond or even instead of leadership. For a long time, leadership was primarily understood in technological and economic terms; more recently also with Sustainable Development Goals or the Twin Transition in mind. This, however, will not be sufficient in the future. We point to a number of complementary and alternative normative ambitions beyond various types of leadership, which might be more appropriate in view of the challenges associated with possible future disruptions. Current debates in EU R&I policy: Are we on the right track?

Against the backdrop of this framing, the paper will revisit the main elements of the 2nd Strategic Plan of Horizon Europe (European Commission 2024) in order to assess to which extent they fit within the frame of the three overarching strategies, or at least take the issues raised by disruptive developments into account in the definition of priorities and instruments ("future proofing"). Depending on the state of the current debate about FP 10, we will also revisit key elements from recently published position paper on the next framework programme for research and innovation; also with the intention to discuss whether they take the need for coping with potentially disruptive changes into account.

REFERENCES:

- Wanzenbˆck, I., Wesseling, J. H., Frenken, K., Hekkert, M. P. & Weber, K. M. (2020). A framework for mission-oriented innovation policy: Alternative pathways through the problem–solution space. In: Science and Public Policy 47(4), 474–489.
- Weber, M., Wasserbacher, D., Kastrinos, N. (2023): Foresight on Demand: Foresight towards the 2nd Strategic Plan for Horizon Europe, Brussels, European Commission
- European Commission (2024): Horizon Europe. Strategic Plan 2025-2027, Brussels, European Commission

The Emergence and Consolidation of Evaluation Systems

Vitalii Gryga, Olha Krasovska, Yuliia Ryzhkova
InstituteforEconomicsand ForecastingNAS of Ukraine, Ukraine
Ukrainian Evaluation Association, Ukraine
Institute for Economics and Forecasting NAS of Ukraine, Ukraine

EVALUATION OF R&D AND INNOVATION POLICY IN UKRAINE:MISSING ELEMENTS

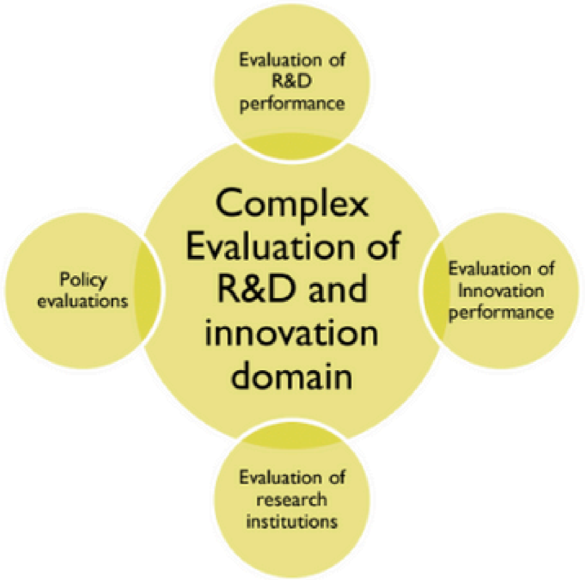
ABSTRACT

Although the full-scale war is still ongoing, discussions on the different modalities of the recovery process are quite active in Ukraine. A series of recovery conferences happened after the full-scale invasion brought to the agenda the necessity to have a clear vision of the multi-optional recovery strategies.The Ukraine Facility Plan, serving as the basis for the implementation of the EU's financial support program for Ukraine in 2024-2027, replaced the Ukraine Recovery Plan, which was rather a collection of poorly linked ideas and proposals than a strategic document.

The paper discusses the complex approach suggested to the evaluation of R&D and innovation domains in Ukraine aimed at designing evidence-based policy making. It is suggested to consider four main elements that build the complex evaluation approach (see figure 1):

1. Evaluation of R&D performance.
2. Evaluation of innovation performance.
3. Evaluation of research institutions.
4. Policy evaluations.

FIGURE 1. BUILDING ELEMENTS OF THE COMPLEX EVALUATION OF R&D AND INNOVATION DOMAIN



Source: developed by authors

Conducting proper evaluations of all mentioned structural parts will ensure complex evaluation of R&D and innovation domains and thus contribute to the building of evidence-based policy fuelling progressive economic development.

We suggest approaching evaluation of R&D performance through the model based on the correlation between the dynamics of scientific personnel and the scientific intensity of GDP in the long-term perspective. Results are based on panel data, which consists of a heterogeneous sample of countries, confirming the existence of a direct proportionate relationship between the dynamics of GDP science intensity and the workforce. Our research proves the validity of the proposed model with a high level of probability. The panel data reveals different patterns in S&T development. Some countries demonstrate low elasticity of research personnel intensity, while the elasticity for other in countries are quite high. An important point is that elasticity is not constant and varies accordingly to the level of S&T development and countries' policy approaches to S&T and innovation development.

As for Ukraine, the relation between R&D funding and research personal intensity is pretty straightforward (see picture 2) over 2010-2020. Moreover, in 2016-2020, the response of research staff on the decrease of R&D funding became stronger.

Picture 2. Comparison of indicators of GDP science intensity and the workforce in Ukraine.

Source: Own calculations based on data from the State Statistics Service. URL:https://ukrstat.gov.ua/

The practical value of the conducted analysis lies in justifying target indicators to be considered during the development of the strategic documents and key targets. The model allowed us to calculate the necessary level of R&D funding in Ukraine by 2030 to achieve at least 40% of the EU-27 level, which was in 2021.

Innovation performance, in turn, is suggested to be evaluated using the composite indices.

The relevance of the approach sparks debate among scholars, given that innovations are inherently unpredictable. In addition, composite indices are static," meaning they do not consider the innovation process, which can be stretched over time, while the situation is changing very fast,and policy should respond promptly.

Comparability of indicators remains another significant challenge because although the indicators themselves are created for comparison, the data and procedures used for their collection and interpretation are not identical across all countries nor unified across all fields of science, technology, and research.

Furthermore, the relative importance of individual factors, the relevance of input data, the causal relationship between input and output data, as well as the frameworks and conditions for innovation creation are barely considered during the indices development. The gap between investment and results is particularly unclear and under-researched: investments in innovation cannot be traced to specific outcomes, and the causal relationship, notably, diminishes over time. Thus, indicators do not reflect the time lag between investments in innovative activities (input) and output. The time lag is not only unknown, but it is likely to vary across different types of innovative activities.

Despite the mentioned controversial issues, and even if indicators at best can only identify strengths and weaknesses rather than explain them, composite indices provide a general overview of the innovation system of a given country and thus could be considered as a relevant approach to the evaluation of innovation performance in a long-term period. While from a short-term perspective, most countries with an underdeveloped innovation ecosystem require the implementation of other approaches for evaluation, e.g., specific surveys.

There is no unified approach to the evaluation of research institutions in Ukraine yet. Thus, the assessment of the effectiveness of research institutions of the National Academy of Sciences of Ukraine is being conducted according to the prospective methodology renewed in 2023. The methodology takes into account the experience of applying procedures and criteria for evaluating the activities of scientific institutions in Germany, Austria, the United Kingdom, Poland, the Czech Republic, and soon. Based on the results of the evaluation, the research institute is entitled to the category depending on what future steps are defined (‘green’ light to continue operations, enhance international cooperation, close/reorganise the institution). At the same time, there is a methodology for a state certification of scientific entities under revision.

Recently, the Ministry of Education and Science of Ukraine has developed a new methodology for assessing the effectiveness of research institutions and universities. The methodology is currently under public discussion, which will last until May 17, 2024. It aims to allow the conduct evaluation of scientific work according to new criteria, including consideration of research contributing to the development of global science, economy, defence capabilities, and Ukrainian society, as well as in line with the principles of open science. It is planned that evaluation will be conducted simultaneously for all research and higher education institutions within specific scientific fields. The evaluation process will take place in digital format through the National Electronic Scientific Information System, ensuring transparency and efficiency. So, it's too early to conclude on the relevance of the new methodology as well as to report on the unified approach of the evaluation of research institutions.

Policy evaluations are the most challenging element of the complex evaluation of R&D and innovation domains because of:

- poor coordination of policy documents in the R&D and innovation domain;
- low enhancement of the mentioned policy documents;
- constant underfinancing of the implementation;
- absence of the independent evaluations of the policy documents and policy implementation.

The last cause is the most crucial in ensuring evidence-based policymaking as the new policy cycle starts without proper consideration of effectiveness and lessons learnt from the previous one. It is worth noting that this is a common problem for policymaking in Ukraine. Although the majority of the policy documents include certain indicators of their implementation, there is no culture to conduct independent evaluations. This, we consider policy evaluations as the weakest element of the complex evaluation of R&D and innovation domain.

Ukraine demonstrates good potential in ensuring two out of four elements of the complex evaluation of R&D and innovation domain, namely evaluation of R&D and innovation performance. The attempt to unify evaluation of research institutions performance has been undertaken recently, while it's too early to assess the relevance of the approach. Policy evaluations remain the weakest element of the complex evaluation of R&D and innovation domains because of the absence of an independent evaluation culture.

REFERENCES:

- Granger,R.C.(2020)InnovationExplained.Leicester:InnoHouse.h
- https://files.nas.gov.ua/text/Documents/Metodika_ocinuvanya_2023.pdf

Nima E. Groji
Technical University Dublin, Ireland

ADVANCING MONITORING AND EVALUATION IN TRANSFORMATIVE INNOVATION POLICY:INSIGHTS FROM EU COHESION PROJECTS

ABSTRACT

Transformative Innovation Policy (TIP) stands as a critical mechanism for addressing contemporary socio-economic and environmental challenges, particularly within the European Union (EU). Amidst the imperative of achieving sustainability objectives and combating climate change, TIP initiatives are essential for driving systemic change. Central to supporting such initiatives are EU Cohesion policy funds, which serve as primary instruments for financing innovation and fostering a fair transition. However, the effectiveness of TIP interventions necessitates robust monitoring and evaluation (M&E) frameworks to assess their impact and guide future policy formulation. This paper aims to contribute to the enhancement of M&E practices in the context of TIP by proposing a new conceptual framework and conducting an empirical analysis of EU Cohesion policy evaluations.

Transformative Innovation Policy (TIP) plays a pivotal role in addressing contemporary global challenges, particularly concerning climate change and sustainable development goals. Within the European Union (EU), cohesion policy funds serve as crucial financial instruments for fostering innovation and facilitating a just transition. This paper focuses on the monitoring and evaluation (M&E) of TIP initiatives. It begins by examining the varying levels of sophistication in M&E processes, ranging from basic recipient satisfaction surveys to advanced causal econometric analyses. Subsequently, it surveys causal inference techniques, emphasising their importance in robust evaluation frameworks. Analysis of recent EU Cohesion project evaluations reveals a limited adoption of causal inference techniques, particularly within innovation and environmental-related programs. Identified gaps in M&E practices prompt recommendations for enhancing evaluation methodologies, advocating for the integration of real-time data, and contrasting traditional M&E approaches with modern ones. Our analysis provides an overview of the conceptual underpinnings of transformative innovation policy for sustainability. It elucidates the significance of TIP in addressing complex societal challenges and achieving long-term sustainability objectives within the EU context.

Here, the survey methodologies employed in monitoring and evaluating TIP initiatives have a focus on the varying degrees of sophistication delineated by Storey's "six steps to heaven" scale. Emphasis is placed on advanced evaluation techniques, particularly those at the sixth step of the scale, such as ascausal inference methodologies. The analysis of data delves into the analysis of M&E exercises within the context of policy interventions,

exploring the design and implementation of evaluation frameworks. Attention is given to identifying areas for improvement and gaps in current M&E practices, particularly concerning the evaluation of innovation and environmental-related programs. After this analysis, we concluded with recommendations for enhancing M&E practices in the realm of TIP. Suggestions include the adoption of real-time data integration, the juxtaposition of traditional and modern M&E approaches, and the promotion of causal inference techniques for robust evaluation frameworks. The evaluation of programs typically involves three main types: ex-ante evaluation, mid-term evaluation, and ex-postevaluation. Ex-ante evaluation occurs before program implementation, ensuring relevance and coherence. Mid-term evaluation happens during implementation, providing insights into progress. Ex-post evaluation takes place post-implementation, assessing goal achievement, efficiency, and sustainability. The proposed evaluation framework builds upon traditional approaches, incorporating additional steps for a more continuous process. It aims to align with the evaluation of investment projects and monitoring processes. Starting with defining expected impacts, the framework proceeds to program/policy design and input identification. Methodologically, it aims to meet Storey's step 6 requirements by enabling the estimation of policy effects through counterfactual methods and quantification of impacts.

Our analysis reveals that the utilisation of advanced techniques, such as counterfactual impact evaluation methods, has been limited in Cohesion Policy evaluations conducted by EU members since 2015. This trend is particularly pronounced when evaluations target thematic objectives related to environmentally sustainable dimensions of policy interventions. From our findings, we draw several recommendations for enhancing monitoring and evaluation (M&E) activities within the realm of transformative innovation policy (TIP) in the contemporary context. Firstly, evaluation efforts should not be confined to the conclusion of a policy program but should occur at all stages, informing goalsetting from the outset. Continuous monitoring throughout the program's duration, facilitated by publicly accessible dashboards updated with real-time data, is essential. Data collection should encompass seemingly inconsequential variables, including information on non-recipients of policy support, to establish a counterfactual control group for comparison. Additionally, tracking outcome variables not officially designated as policy objectives can help identify unintended indirect effects of the policy.

While our paper provides valuable insights, it is not without limitations. We acknowledge its focus solely on the M&E dimension of TIP without fully encompassing all dimensions of the subject. Nonetheless, we aim for this paper to enrich discussions on monitoring and evaluation in the realm of TIP among innovation scholars and offer practical guidance to policymakers tasked with implementing M&E practices for real-world innovation policies in times demanding urgent change. In summary, this paper advocates for the refinement of M&E practices in transformative innovation policy. By proposing a new conceptual framework and conducting an empirical analysis of EU Cohesion policy evaluations, it seeks to contribute to the advancement of TIP interventions and the achievement of sustainable development goals within the European Union.

REFERENCES:

- Trischler, J., Svensson, P.O., Williams, H., Wikström, F. (2022). Citizens as an innovation source in sustainability transitions–linking the directionality of innovations with the locus ofproblem in transformative innovation policy. PublicManagement Review.
- Esmail, B.A., Geneletti, D. (2018). Multi-criteria decision analysis for nature conservation: a review of 20 years ofapplications. Methods in Ecology and Evolution, 9, 42–53.
- Bjerke, M.B., Renger, R. (2017). Being smart about writing SMART objectives. Evaluation and Program Planning, 61, 125–127

Volpini Giulia, Joly Pierre-Benoit, Matt Mireille
INRAE Toulouse-Occitanie, France

FORMATIVE AND IMPACT-ORIENTED EVALUATION AS A PROJECT MANAGEMENT TOOL: FROM KPI TO KSD

ABSTRACT

Transformative innovation projects that attempt to address the key challenges that our society is currently facing have increased rapidly over the last decade as a solution to wicked problems, including climate change. These projects, focused on societal and/or ecological challenges, involve many different actors, such as social, grassroots, research and civil society innovators (Molas-Gallart et al., 2021; Todorović et al., 2015). Monitoring and guiding these projects to ensure they achieve their objectives, contributing to big societal challenges, is therefore a core issue. Nowadays, in management science, strategic decision-making is considered the key to solving this problem. Strategic decisions, which aim to align a project with its changing environment, can be implemented through performance management, a strategic process fostering the viability and success of organisations such as transformative innovation projects (Dolence & Norris, 1994; Patrick, 2013). The main tool of this approach is KPI (Key Performance Indicators). They are a popular and widespread method of measuring the performance of an organisation by focusing on the key factors that contribute to the realisation of its outputs. They provide quantitative information regarding the performance of the organisation, while showing institutional progress through milestones and indicators. They aim to provide core information to the decision-makers to make strategic moves, in an attempt to enhance profitability and longterm success. Moreover, KPIs should allow an organisation to learn and improve continuously. In essence, KIPs promise to guide the organization, to make it more effective and more competitive (Beatham et al., 2004; Patrick, 2013). In this article, we oppose to this mainstream approach based on managers' responsibility and effectiveness a formative impact-oriented approach to collective project management. The construction of this monitoring approach is the result of two years of co-experimentation of the ASIRPArt methodology with several French transformative innovation projects willing to contribute to agroecological transition. ASIRPArt is a formative assessment methodology that aims to support research and development projects to amplify their impacts towards desired societal futures and navigate uncertainty. Its objective is to guide the innovation process to determine whether expected transformations are taking place and to adjust iteratively ongoing experiments, while enhancing actors' learning in a continuous process (Bhat, 2019; Joly et al., 2019; Matt et al., 2023). But, how to use formative and impact-oriented evaluation methodology as a project management and monitoring tool for transformative innovation projects?

In this context, formative evaluation is considered a form of reflexive practice aiming at helping project members address their transformative objective and contributing to their ability to do so. Difficulties are therefore seen as a learning opportunity on the context, conditions, and activities leading to transformation processes. In contrast to result-oriented and performanceoriented approaches which focus on accountability and efficiency, reflexive monitoring and evaluation aim to build the structures for discussion of values and visions (Molas-Gallart et al., 2021).

Moreover, this approach is based on several characteristics:

- Impact-oriented visions, as the projects aim to contribute to a societal transformation, not just to be efficient.
- Iteration and adaptability, as when addressing wicked problems, system change can be complex and unpredictable. Therefore, the project actions need to adapt and change.
- Participation, as the inclusivity characterizing the projects should also be present in the monitoring process. Moreover, formative evaluation aims to facilitate participation and open debate, channelling the eventual conflicts and differences in interest and perceptions.

When talking about formative evaluation as a project management tool, we distinguished two phases: 1) the phase of design and co-construction of the process and 2) the phase of actual monitoring of the project activities. Over the first phase, project members are asked to define the desired transformations to which the project aims to contribute (step 1) and to analyse the ecosystem of the project, namely the societal context in which the project is immersed and that creates uncertainty for the realisation of the desired transformations (step2). Finally, they have to define the actions that will contribute to those transformations, taking into consideration the ecosystem and tracing the logical process that connects their actions to the transformations (step 3). Once this exercise has set the first draft of the project plan, the second phase of monitoring is fundamental to allow the project members to follow the advancements according to their transformative objective and to rediscuss the project activities when they deviate from the target. In order to do so, a few tools can be useful:

- Impact pathway (IP): it describes the non-linear process of how the knowledge and the actions mobilised by the LL turn into outputs, and how these outputs eventually translate into societal transformations and impacts through the intermediary ecosystem. Thanks to this tool, the projects can create a graphical representation of the required steps to create impact.
- Impact narrative: it allows explicating the dynamics schematised in the IP.

- Key Steering Descriptors (KSD): they correspond to the tipping points of the project, namely those events (external or internal to the project) that condition the realisation of the desired transformation. Moreover, they are defined as descriptors as they define the characteristics of an action and do not have a common reference, as in the case of indicators. Therefore, they can be qualitative or quantitative. The KSD adapt to the aims and activities of the project and are based on the project plan and objectives. They are thought to support project members in decision-making and the actual steering of the project

The methodology, as well as the KSD, remain flexible and mutable according to the learning process over the progress of the project, according to the principles of adaptive management. To help the project navigate uncertainty, these tools should be mobilised regularly to monitor whether the project is still going in the direction of the desired transformation. According to the principle of participation, the monitoring should involve all project members, as each of them would be responsible for the KSD concerning their specific activity and contribution to the project mission. The co-experimentation with the French transformative innovation projects allowed us to test the methodology in different forms to attain this final version. We confronted the methodology with the actual difficulties and ambitions of steering this kind of project, which brought the approach to be simplified and perfectioned.
In conclusion, the proposition of KSDs challenges the KPIs in project management. We oppose a qualitative approach, which aims to be more inclusive of the monitoring needs of projects, to a quantitative and reductive one. The choice of using descriptors instead of indicators reflects this same issue while integrating core aspects such as flexibility, adaptation and iteration of the tools. Moreover, the ASIRPArt monitoring methodology promotes an impact-oriented approach to a performance-oriented one, to better answer the reason for being of transformative innovation projects to transform society. Finally, this approach fosters a collective responsibility toward project management, contrasting the widespread tendency of considering managers as the main and only responsible for the good realisation of the project.

REFERENCES:

- Beatham, S., Anumba, C., Thorpe, T., & Hedges, I. (2004). KPIs: A critical appraisal of their use in construction. Benchmarking: An International Journal, 11(1), 93-117. https://doi.org/10.1108/14635770410520320
- Bhat, B. (2019). Formative and Summative Evaluation Techniques for Improvement of Learning Process. 07(05).

- Dolence, M. G., & Norris, D. M. (1994). Using key performance indicators to drive strategic decision making. New Directions for Institutional Research, 1994(82), 63-80. https://doi.org/10.1002/ir.37019948207
- Joly, P.-B., Matt, M., & Robinson, D. K. R. (2019). Research Impact Assessment: From ex post to real-time assessment. fteval - Platform for Research and Technology Policy Evaluation. https://doi.org/10.22163/fteval.2019.326
- Matt, M., Robinson, D. K. R., Joly, P.-B., Van Dis, R., & Colinet, L. (2023). ASIRPARealTime in the making or how to empower researchers to steer research towards desired societal goals. Research Evaluation, 32(2), 412-425.
- Molas-Gallart, J., Boni, A., Giachi, S., & Schot, J. (2021). A formative approach to the evaluation of Transformative Innovation Policies. Research Evaluation, rvab016. https://doi.org/10.1093/reseval/rvab016
- Patrick, J. (2013). Key Performance Indicators: Advances in their Development and Evaluation. https://hdl.handle.net/102.100.100/535132
- Todorović, M. Lj., Petrović, D. Č., Mihić, M. M., Obradović, V. Lj., & Bushuyev, S. D. (2015). Project success analysis framework: A knowledge-based approach in project management. International Journal of Project Management, 33(4), 772-783. https://doi.org/10.1016/j.ijproman.2014.10.009

Aura Kivilaakso, Johanna Kolhinen
University of Helsinki, Finland

RESEARCH ASSESSMENT AS A TOOL FOR STRATEGIC MANAGEMENT

ABSTRACT

The University of Helsinki (UH) is the oldest and largest institution of academic education in Finland, an international scientific community of 40,000 students and researchers. It operates on four campuses in Helsinki and in 10 other localities in Finland. The University accommodates 11 faculties, several independent research-oriented institutes, multidisciplinary research networks and campus units, as well as units attending to duties of a national authority. Through the power of science, the University has contributed to society, education, and welfare since 1640. Its strategic plan currently emphasises four research themes: wellbeing for humans and our environment, a humane and fair world, a sustainable future for our planet, and the possibilities that limitless curiosity opens – a universe of ideas and opportunities.

Towards enhancement-led approach of research assessment

In 2018–2019, UH renewed its research assessment framework with an enhancement-led approach. The aim of the Research Assessment University of Helsinki (RAUH) was to produce an overview of the quality and impact of the research conducted at the University, to help identify future research opportunities and support the renewal of research. The framework development was based on the idea of continuous development and quality enhancement instead of a one-time-exercise of appraisal. A key driver for the framework development was to support the University's 2021–2030 strategy implementation. At the same time, the comprehensive assessment was to fulfil the national legal requirements and cover all research carried out at the University. In Finland, universities are expected to assess their own research frequently and publish the results openly. However, there is no common framework concentrating on the quality of research, whereas the Quality Audit system for Higher Education Institutions in Finland by FINEEC covers education, research, and outreach elements of the quality systems in more general manner.

The 2018–2019 assessment was built around self-assessment concentrating on three criteria: scientific quality, societal impact and research environment and unit viability. The units were instructed to reflect upon the research and the research environment in a nuanced way to have a truly useful basis for further development. The four external expert panels also valued the unit's capacity for critical selfreflection. The panels were asked to focus on the unit's readiness to deal with possible deficiencies, e.g., by describing already taken or planned actions, rather than the deficiencies per se. In accordance with enhancement-led evaluation, self-evaluation is primarily a tool for improving operations.

The 15 Faculties and independent institutions assessed themselves on the department level, in total, 39 units of assessment took part. In addition, faculty-level summaries were collected.
The self-assessment report was instructed to be structured according to the predesigned headings, but the units were encouraged to freely decide on the use of any sub-headings. In the report, the units were expected to carry out as reflective self-assessment as possible, identify areas in need of development, and provide a concrete description of the operations and results. The first part of the self-assessment report focused on background information. The core of the self-assessment was the second part: the description of the organisation, profile, mission, and goals of the unit. The unit's performance and operations were primarily assessed against those measures. Self-assessment included reflection on the strengths and weaknesses of the described actions. Supporting metric data was provided on funding, personnel, publications, and other outputs as well as on doctoral research. In section three, the units described the self-assessment process.

Making use of the RAUH 2018–2019 results in continuous strategic development
RAUH 2018–2019 offered a solid basis for the long-term development of research within academic units. The assessment results have thus been actively used since the assessment report's publication. Each academic unit has, for example, utilized the development areas identified by the evaluation panel in their annual implementation plans following the strategy of the University. In addition, a more ambitious measure was taken in 2022–2023, when the University and its academic units elaborated 'Roadmaps for Implementing Research Themes' that contribute to the objective defined in the University's strategy: "In 2030, the University will enjoy an increasingly established international standing as a scientific partner, especially thanks to its ground-breaking discipline-specific expertise as well as its multidisciplinary and cross-disciplinary research." Roadmaps for Implementing Research themes was part of the implementation of the UH strategy, but accordingly provided inputs for planning the follow-up of the RAUH 2018–2019, too. The follow-up of the RAUH 2018–2019 took place in 2022–2023 as a mid-term evaluation checkpoint, focusing on the development steps taken by the faculties and independent institutes. The follow-up consisted of 15 faculty-level qualitative self-assessment reports (6–10 pages each) focusing on: 1) Key take-aways from RAUH 2018–2019, 2) Key development steps and actions taken based on the results, and 3) Alignment with the current UH strategic plan. The units benefited from the work they had done to develop their Roadmaps for Implementing Research Themes in writing the RAUH 2018–2019 follow-up self-evaluation reports. All self-assessments were compiled and shared openly across the units. Emphasis was on the quality of operations, and metric data was used only as a background, not as the target of the assessment itself. The self-assessments were delivered to the RAUH 2018–2019 Panel Chairs who wrote their feedback on the progress. A joint discussion seminar for the unit leadership, RMA (research managers and administrators) and other interested

services and university community was organised on 29 March 2023. The Panel Chairs were invited to dialogue with the units.

The RAUH enhancement-led approach from 2018–2019 to research assessment is well aligned with current developments with responsible research and researcher assessment (e.g., CoARA). Thanks to the qualitative approach, we can see how we are progressing with ways of managing and leading research. It allows the University leadership to gain a rich understanding of the development work and its meaning for the units and faculties. The results of the follow-up show a variety of concrete development steps stemming from the panel feedback and development with impact. The quality of operations ensures the quality of results, too

The future of research assessment – challenges of interdisciplinarity

In 2025, the University of Helsinki will undertake the overall assessment of its research again. The positive experience gained from the RAUH 2018–2019, and the Roadmaps for Implementing Research Themes will allow the RAUH 2025 assessment to draw on some of the same elements as the previous assessment. However, our intention is to rethink the whole assessment process from the perspective of the University's management and strategic development. The University of Helsinki's strategic goal is to invest in cross-cutting multidisciplinary and interdisciplinary research. Interdisciplinary research builds on strong disciplinary expertise with an integrative approach. The purpose is to facilitate the collision of different perspectives to generate so-called “landscapes of knowledge”. Through the interdisciplinary cooperation, the University aims to reach novel research directions that seek solutions to major global problems. RAUH 2025 is not just an overall assessment of research, but rather an important next step towards the implementation of the University's strategic goal of promoting interdisciplinarity. In 2025, the statutory research assessment will be harnessed as a tool for strategic management and promoting multi- and interdisciplinary research. RAUH 2025 focuses on the management and state of multi- and interdisciplinary research at the University. Its goals include highlighting strong research areas, identifying emerging research fields, and assessing the University’s success in responding to the societal problems that its research is tackling as outlined in the UH strategy 2021–2030. To allow academic units to monitor the success and progress of their development after the RAUH 2018–2019 and to keep the continuous development cycle as smooth as possible, the RAUH 2018– 2019 criteria will also be applied in the RAUH 2025. However, this time the criteria will be considered as themes to be assessed rather than as assessment criteria in themselves. During autumn 2024, we will analyse in depth the means of practising and measuring research activities, outputs, and outcomes that support multi- and interdisciplinarity and will redefine the criteria for the RAUH 2025. The RAUH 2025 development will be supported by several other concurrent activities, such as the launch of a training program for Vice-Deans for Research, which will focus on strategic research management. The training program will be designed from the basis of the Roadmaps for Implementing Research Themes and focus group discussions between the Vice-Rector for Research and the University’s researchers with the best expertise in interdisciplinary research. Key questions in the focus group discussions are how to measure and manage the success of interdisciplinarity at the University. This way we invite the people assessed to define the success from their own perspective and use the understanding gained towards reflecting on the RAUH 2025 evaluation criteria.

Integrating interdisciplinarity into the overall assessment of research is a new and ambitious attempt in the field of research assessment. To succeed, it requires careful and regenerative planning. During the autumn 2024, we will define detailed criteria for the implementation of RAUH 2025. In our conference presentation in December 2024, we will present our plan for the implementation of RAUH 2025 and for promoting interdisciplinarity as part of the overall assessment of a comprehensive university's research activities

Dr. Jan Stede, Dr. Swaroop Rao, Marc Blauert, Dr. Barbara Schlomann, Julian Schaper
Technopolis Deutschland GmbH, Germany
Fraunhofer Institute for Systems and Innovation Research ISI, Germany
Technopolis Deutschland GmbH, Germany
Fraunhofer Institute for Systems and Innovation Research ISI, Germany
Technopolis Deutschland GmbH, Germany

EVALUATING ELECTRIC CAR SUBSIDIES - AN EX-POST APPROACH LEADING INTO AN EXANTE APPROACH

ABSTRACT

Both at European and national level, the requirements for reporting on policy measures introduced to achieve EU-wide and national energy efficiency and climate targets have increased considerably in recent years. This means that also the ex-post and ex-ante evaluation of these policies is becoming increasingly important. In this paper, we present an evaluation methodology which combines the ex-post and ex-ante evaluation of a climate policy by evaluating the €10bn German electric vehicle subsidy scheme.

In many European countries, the transport sector faces substantial challenges to decarbonise to meet the climate neutrality targets set by the EU and national governments. Transport is the only sector in the EU, where emissions have increased since 1990.4 Public subsidies for the purchase of electric vehicles have been introduced in various countries as a means of incentivising consumers to shift away from fossil-fuel powered personal transport. The electrification of the transport sector is a crucial element in transforming not only the demand side of the mobility sector (in terms of consumer choices), but also the supply side (in terms of impacts on the German and European automobile industry). Manufacturers are compelled to innovate in terms of technology innovation, and policymakers are compelled to innovate in designing suitable policies to ensure the roll-out of decarbonised mobility options such as electric vehicles. Ex-post and ex-ante evaluations of such policies are therefore necessary to better inform policymakers, the industry, and consumers alike of future implications of policies and to improve the data base for the monitoring of energy efficiency and climate targets.

The purchase premium paid by the Federal Government and manufacturers for passenger cars with electric, hybrid and hydrogen / fuel cell drive systems (the so-called “Umweltbonus" / "environmental bonus”) was a key climate policy measure in Germany for the electrification of road transport. With the subsidy program, private households and companies were granted a financial subsidy of up to 9000€ when purchasing or leasing an electrically powered vehicle. Since the introduction of the environmental bonus in 2016 up to the end in 2023, state funding of more than €10 billion has been approved, which has

been used to support the registration of almost 2.2 million electric vehicles.

In this paper, we present the methodology and first results of the ex-post evaluation of the program over the years 2016-2023, which was carried out by the authors of this paper. In addition, we refer to the ex-ante assessment of energy and GHG savings of the program until the year 2030, which was directly based on the results of the ex-post evaluation. 6 The primary objective of the paper is to strengthen the evaluation infrastructure at the disposal of evaluators of transport policy measures by establishing a set of indicators, the data sources to be considered, as well as a standardised and transparent methodology to guide evidence-based policymaking. In particular, the methodological linking of ex-post and ex-ante assessment of energy and GHG savings is intended to improve the data basis for estimating the achievement of short-, medium- and long-term climate targets.

Ex-Post Evaluation: Methodological approach to calculate gross effects and net effects of the subsidy

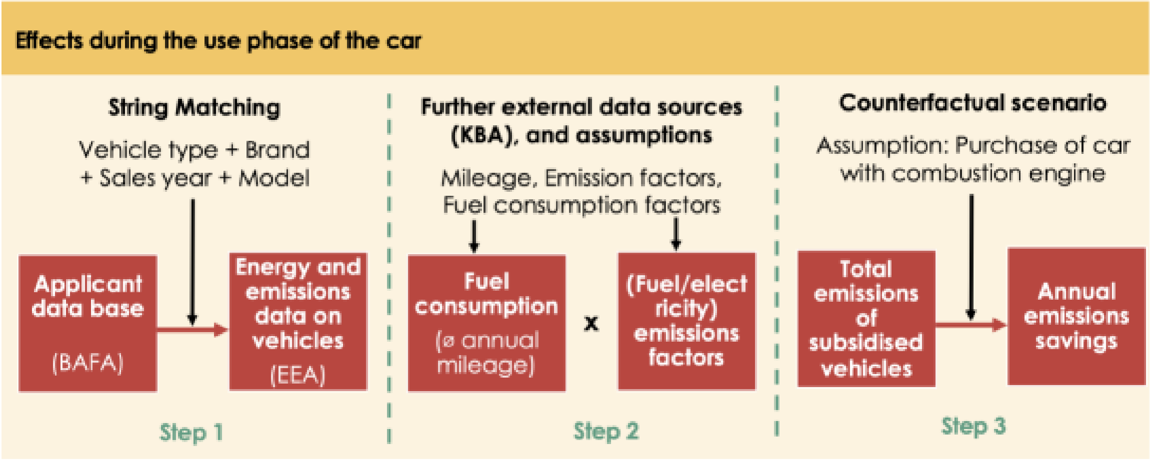
The ex-post evaluation is based largely on the methodological guideline for evaluations of energy efficiency measures created on behalf of the German economics and climate ministry BMWK (Fraunhofer ISI et al., 2020). The aim of this standardized evaluation framework for energy and climate policy measures is to ensure cross evaluation and cross-program comparability. This evaluation follows the specifications and recommendations of the methodological guidelines and adapts them to the specific needs of the electric vehicle subsidy program.

For some of the key indicators (e.g., GHG emissions savings, costs per tonne of avoided emissions), we calculate both gross effects (effects of the policy without considering whether the estimated effects are actually induced by the policy), and net effects. For the calculation of net effects, we address and quantify the additionality of the policy by correcting for free-rider effects (i.e., the vehicle would have also been bought without the policy, anticipation effects (i.e. buying the car earlier than planned as a result of the subsidy), as well as rebound effects.

For the quantification of gross effects, we follow a three-step approach (Figure 1). Our analysis is based on a unique official micro dataset of the 2.2 million cars bought by subsidy beneficiaries from the Federal Office of Economics and Export Control (BAFA), which administered the program.

In a first step, this dataset is matched with other data sources such as the vehicle registration data of the European Environment Agency (EEA), which contains GHG emission and energy consumption data. This linking allows us to determine the exact emission and consumption values of the subsidized vehicles at model level. In a second step, we calculate exact GHG emissions and energy consumption levels for the subsidized vehicles and their respective reference vehicles at individual vehicle level. We base this calculation on assumptions on annual mileage over the lifespan of a vehicle, current GHG

emissions factors for electricity and fuel production, primary energy factors, electric driving shares for PHEV as well as real consumption factors of electricity and fuel for all considered vehicles to account for real life consumption, obtained from the literature (BIT & Fraunhofer ISI, 2021; ICCT, 2022; Öko-Institut et al., 2023). In a third step, to estimate gross effects at the vehicle level, we compare energy consumption and GHG emissions of calculate exact GHG emissions and energy consumption levels for the subsidized vehicles and their respective reference vehicles at individual vehicle level. a subsidized vehicle to a (hypothetical) fossil 'reference vehicle' for all subsidized vehicles. Our key identifying assumption is thus that without the subsidy, a vehicle with an internal combustion engine in the same vehicle class would have been bought. We provide support for this identifying assumption in the paper.
FIGURE 1 OVERVIEW OF THE CALCULATION OF GROSS GHG EMISSION EFFECTS



In order to compute net effects, we quantify free-rider and anticipation effects, as well as the direct rebound effects based on a large-scale representative survey of 2,519 recipients of the subsidy (corresponding to a response rate of 16.9%). The inclusion of free-rider and anticipation effects corrects for the fact that some recipients would have purchased an electrically powered vehicle even without the subsidy. The direct rebound effect accounts for increasing distance travelled relative to the formerly driven internal combustion engine vehicle. Our results show that gross and net effects can differ significantly, and underlines the importance of including net impact monitoring when calculating the effects of energy and climate policy instruments.
The results of the ex-post evaluation provide insight into the efficiency of the policy in terms of the amount of funding used to encourage additional electric vehicle registrations, as well as associated savings of GHG emissions and energy. The final evaluation results are not published at the moment, but will be publicly available by the time the conference takes place. Preliminary results of the evaluation show that since its inception in 2016 the subsidy supported the purchase of 1.8 million electric vehicles until the end of 2022 and contributed to savings of 2.72 Mt CO2e in net GHG Emissions savings, as well as 10,910 GWh

of net primary energy savings. The estimated rebound effects were around 2% to 3%, and the aggregated free-rider and anticipation effects (combined) ranged from 34% to 46%, depending on the year.

Ex-ante evaluation

The indicators calculated and obtained as part of the ex-post evaluation are then used as the basis for the exante estimations of the subsidy's impact. The methodological basis for these calculations was laid down in guidance for the ex-ante assessment of climate policies (Schlomann et al. 2022), which was directly based on the methodological guideline for the ex-post evaluation. We estimate for the years until 2030 the projected development of the core indicators of our study, namely the specific GHG emissions savings, the increase in demand for E-Mobility (i.e. the number of EVs on the streets), the free-rider and rebound effects, as well as the subsidy efficiency (cost per tonne of CO2e avoided). This coincides with the reporting requirements of the National Energy and Climate Action Plans (NECPs) 2021-2030 under the EU governance regulation8 . These estimations are undertaken using the following relationship: Absolute Savings = specific savings x activity size. The absolute savings can be of GHG emissions, or energy consumption. The activity size refers to, for instance, the amount of funding made available for the subsidy, or the number of additional electric vehicles on the roads.

The ex-ante estimations can be static or dynamic. In the static analysis, the current decision-making situation is always taken into account and considers the savings that would be achieved with a measure if it were to be continued in the current decision-making situation. This approach is intended to take into account past implementation efficiencies and other parameters relevant for the ex-ante estimation. In the dynamic view, activities can also be taken into account if a decision has not yet been made, but an intention has already been formulated.

Given that the "environmental bonus" subsidy programme came to an abrupt stop in 2023, it is unlikely that the same subsidy model will be continued by the German Ministry of the Economy and the Climate, but given the substantial lack of progress in meeting the transport sector decarbonisation targets, it is also likely that there will be measures taken to support E-Mobility for both households and organisations. A dynamic approach for the exante estimations is therefore preferred, to take into account potential changes in the broader economy as well as advancements in technology.

REFERENCES:

- BIT & Fraunhofer ISI (2021). Pkw-Antriebe für die Zukunft: Ökonomische, ökologische und technische Effizienz im Vergleich. 10.24406/publica-266.
- BMWK (2023) Umweltbonus endet mit Ablauf des 17. Dezember 2023. Available at: https://www.bmwk.de/Redaktion/DE/Pressemitteilungen/2023/12/20231216-umweltbonus-endet-mit-ablauf-des17-dezember-2023.html.

- Fraunhofer ISI et al. (2020): Methodikleitfaden für Evaluationen von Energieeffizienzmaßnahmen des BMWi. https://publica-rest.fraunhofer.de/server/api/core/bitstreams/37161a8d-8b1f-4a69-a4c1-98f2372554d5/content.
- ICCT (2022). White Paper - Real-World Usage of Plug-In-Hybrid Vehicles in Europe. A 2022 Update on fuel consumption, electric driving, and CO2 emissions. https://theicct.org/publication/real-world-phev-use-jun22/.
- Öko-Institut e.V., Fraunhofer ISI, IREES, Thünen-Institut (2023): Kernindikatoren des Projektionsberichtes 2023. https://www.umweltbundesamt.de/publikationen/projektionsbericht-2023-fuer-deutschland.
- Schlomann, B., Brunzema, I., Kemmler, A., Bürger, V., Mendelevitch, R. (2022): Methodikpapier zur ex-ante Abschätzung der Energie- und THG-Minderungswirkung von energie- und klimaschutzpolitischen Maßnahmen. https://www.isi.fraunhofer.de/content/dam/isi/dokumente/cce/2022/Kurzpapier_Ex-anteMa%C3%9Fnahmenbewertung_220712_ver%C3%B6ffentlichte-Version.pdf

Lena Tsipouri, Sofia Liarti, Dimitris Pappas, Nikos Maroulis, Haris Papageorgiou
National and Kapodistrian University of Athens; OPIX PC, Greece
OPIX PC , Greece
OPIX PC , Greece
Technopolis Group , Belgium
OPIX PC , Greece

EVALUATING THE RTDI SYSTEMS OF WIDENING COUNTRIES

ABSTRACT

The objective of this paper is to highlight how unique characteristics of expanding economies impact the development of evaluation methodologies for assessing their Research, Technology, and Development (RTD) systems and policies. While the Better Regulation guidelines are typically applied universally, we argue that specific traits necessitate special attention, while in other cases, traditional indicators yield minimal changes or insignificant variations that are unlikely to generate meaningful conclusions.

The differentiating features of the RTD systems of Widening countries

Our hypotheses are that Widening Countries with low performance in R&I face common problems such as:

- Differentiated evolution of R&D versus innovation; the former progresses faster than the latter and this has significant repercussions on the rationale of R&D funding.
- Brain Drain is a common issue in many Widening countries, where highly skilled researchers and professionals migrate to more developed countries in search of better research opportunities.
- Institutional arrangements lead to path dependentt pathogeneses:

1. Systematic lags in all stages of the policy cycles and
2. Reluctance to specialise measures by concentrating funding in specific sectors or missions or clusters
3. Reluctance to address individual challenges with specific measures (following Tinbergen’s recommendation) and using traditional generic schemes to cover all needs

Results from the evaluation of the Greek RTD support 2007-2014

Since 2008, Greece has advanced from being classified as a Catching-up country on the European Innovation Scoreboard (EIS) to a Moderate Innovator, achieving a performance level of 79.5% of the EU average in 20234. While Greece's overall innovation performance still trails behind the EU average, there is a positive trend indicating a reduction in the performance gap vis-‡-vis the EU, surpassing the EU's rate of improvement by 8.5

percentage points.
In 2022-2023, we were members of a consortium which evaluated the entire Programme composed of 18 RTDI Actions designed and funded by the General Secretariat for Research and Innovation (GSRI). The study aimed at assessing the impact of the mix of Actions on both funded and research and production system and the economy, at evaluating the management by the GSRT, at identifying European good practices, and at drawing insightful conclusions to enhance future policy mixes and management of actions and propose recommendations for the next programming period.
We used an intervention logic for each Action, collected data for publications, citations and patents from open repositories using Ai tools and wherever possible used counterfactual analyses.
Comparing R&D versus innovation improvements The funding was too limited to make a large difference but there are several results R&D indicating both quantitative and qualitative progress in terms of publications, citations and the share of high quality publications. Conversely, in terms of innovation measuring patents did not generate any statistically significant results, mainly because the number of patent grants is marginal anyway, hence differences risk being misinterpreted. During the evaluation of each action, we observed a lack of prioritisation in priority areas. Based on our analysis, every project was distributed to at least one SDG and one S3 priority area according to their proposal title. While some sectors and priority areas were more funded than others the concentration was at FoS level 1with no efforts to reinforce sub-sectors ate FoS 3 or 4, nor clusters (with an exception with really marginal funding), or specific types of companies. The bulk of the funding (limited because of the crisis anyway) was concentrated in traditional schemes where all types of companies and disciplines were eligible.
Brain Drain: A Policy measure “Support Postdoctoral Researchers” aimed i.a. to attract highly skilled and educated researchers that left the country. Our team developed an algorithm to track researchers' mobility based on their affiliation (as extracted from their publications), enabling us to categorise them into the locational pattern groups:

- People who were in Greece, when they received the grant, left the country and stayed abroad until now
- People who were in Greece, left and the country but came back
- People who were abroad at the moment they received the grant and came back to Greece at some point after the grant
- People who have never been abroad
- People who were abroad, came to Greece and left again

- People who never came back to Greece Based on our results, the majority, almost half of the grant holders, were and remained in Greece. Out of those who were abroad, 20 individuals have returned to Greece. However, 23 postdoctoral

Based on our results, the majority, almost half of the grant holders, were and remained in Greece. Out of those who were abroad, 20 individuals have returned to Greece. However, 23 postdoctoral researchers who were part of the Greek research system left the country during or after the grant. The 12 researchers who appear to have never come back to Greece etained their original affiliation in their publications.

TABLE 1: RESEARCHER'S MOBILITY FOR THE ACTION "SUPPORT POSTDOCTORAL RESEARCHERS"

Researchers' mobility	Number
People who were in Greece, left abroad and stayed there	23
People who were in Greece, left abroad and came back	15
People who were abroad and came to Greece	20
People who have never been abroad	70
People who were abroad, came to Greece and went abroad again	12
People who never came to Greece	12

Institutional aspects: A common problem in all Actions was the significant delays in proposals' evaluation, contract signing, and payments. The overall average from the submission date until the inclusion was 15.6 months, but it reached up to 25 months in specific Actions. This timing was excessively long and in a particular case led the best applicants to refrain from the grant. Counterfactual analysis showed that they were subsequently the most successful researchers and were partly those who migrated.

Evaluating RTDI initiatives in Widening countries requires adherence to the general Better Guidelines recommendations, while also incorporating additional aspects relevant to the path-dependence of moderate innovators. Complementing standard indicators with such considerations is crucial. AI-tools can play a pivotal role in enhancing this complementarity.

Sanja Vojvodić
Faculty of Political Science, University of Belgrade, Serbia

EVALUATING ENVIRONMENTAL SUSTAINABILITY THROUGH THE LENS OF POSITIVE PEACE

ABSTRACT

Balkan Peace Index encompasses six WB states (Albania, Bosnia and Herzegovina, Croatia, Montenegro, North Macedonia, and Serbia, and the territory of Kosovo4) and covers seven domains (with each of them having three or more indicators) with infographic visualization. On the side of negative peace are two domains (Political Violence and Fighting Crime), while on the side of positive peace, there are five more domains (Regional and International Relations, State Capacity, Environmental Sustainability, Political Pluralism, and Socio-Economic Development). It is designed as a source of data for academic research on peace and security issues in the WB, an e-learning tool for higher education in the WB region, and a mechanism for informing state institutions and international governmental organisations and agencies (UN, EU, OSCE, etc.). It works as a service platform offering public policy proposals, conflict evaluation/monitoring reports and risk assessment/early warning reports to all interested stakeholders. As such, the BPI represents a newly developed culture of evaluation that links international indices with locally monitored and evaluated key political events, including official and non-governmental organization reports related to the domains. A particularly challenging aspect of the evaluation was the BPI domain of environmental sustainability. The escalating environmental crisis poses heightened threats to global security and peace, particularly in nations already grappling with instability. Observable metrics of insecurity, for instance conflict frequencies, prevalence of hunger, and military expenditure, are on the rise, concurrent with indicators of environmental degradation caused by antropogenic climate change and mismanagement of commons. The amalgamation of these security and environmental predicaments triggers compound, cascading, emergent, systemic, and existential risks (SIPRI, 2022). The influence of ecological degradation on conflict is underscored by the significant correlation observed between nations exhibiting the highest levels of conflict, as assessed by the Global Peace Index (GPI), and those experiencing the most severe ecological deterioration. Against this backdrop, BPI acknowledges the inextricable link between environmental sustainability and positive peace construed as the optimal environment for human and non-human "potential to flourish". However, the process of defining environmental sustainability for the Index was characterized by numerous revisions due to its status as a heavily disputed concept within the social sciences. The way out of the dilemma was a consensus among the researchers contributing to the Index to rely on the context-specific knowledge regarding the

environmental sustainability in connection with positive peace in the Western Balkans. Differently put, the aim was to identify and prioritize environmental sustainability issues that possess the greatest capacity to undermine peace. These issues encompassed an assessment of the present condition of natural resources, air quality, and energy systems (given their predominant influence on air quality). Subsequently, the final selection of indicators served to refine the definition of environmental sustainability and provided guidance for the comprehensive monitoring and evaluation process. The qualitative evaluation of indicators was based on the following sources: first indicator – Species protection Index (Map of Life), Wastewater Treatment (UN SDG: 6.3.1), Ecological Threat Report, Tree cover loss indicator (Environmental Performance Index), European Commission Country Insights on the Green Agenda and Sustainable Connectivity, WB-based NGO reports, State audit institutions' reports, and news outlets; second indicator – Outdoor Air Pollution indicator (Social Progress Index), Greenhouse Gas Emissions per capita indicator (Environmental Performance Index), media outlets, reports from environmental protection agencies in the region as well as local NGOs; third indicator – World Energy Trilemma (Energy Security and Energy Sustainability indicators), SDG 7.1.2 (Usage of clean fuels and technology for cooking), SDG 7.2.1 (Renewable energy share in the total final energy consumption), civil society reports (e.g. RES Foundation), EU Energy Community reports, and news outlet. The evaluation also includes key events in the form of social unrest, protests, or wider citizen dissatisfaction that occurred in the observed year and refer to three indicators.

The Western Balkans enclose many intertwined habitats that know no political borders. If one country has a low capacity to protect its ecosystems, negative externalities will be felt in another. The region confronts a substantial risk to environmental sustainability, prompting the author to assess this domain of the BPI as poor, both in its overall evaluation and across individual indicators.

Natural resources resilience grapples with rapid weather shifts from flash floods to lengthier periods of dry conditions, exacerbated by critically low wastewater treatment. Lake Prespa, one of the oldest lakes in Europe, epitomises these dire dynamics in 2023. Despite efforts acknowledged by international indices on species protection, mismanagement of old forests (e.g. Fruška gora in Serbia) significantly decreases biodiversity levels. There are protests against announced infrastructure that might aggravate these frailties across the region, for instance, in Montenegro against undersea exploration of fossil fuels and small hydropower plant constructions in North Macedonia, Bosnia and Hercegovina, and Serbia.

Second, air quality is at an all-time low, and citizens of the Western Balkans continue being exposed to severely unhealthy air quite more than citizens of other parts of Europe. Due to over-reliance on outdated, anachronistic coal industry and inefficient individual combustion plants, more than 30,000 people die prematurely in the region. The air quality context is daunting in Serbia, Bosnia and Herzegovina, North Macedonia, and Montenegro, where political elites avoid discussions regarding pollution. During the last two years, pollution increased for all three regulated pollutants - sulphur dioxide, PM particles, and nitrogen oxides. Also, five years have passed since the implementation of air pollution regulation standards (that were put in place in compliance with the Treaty on the Establishment of the Energy Community on January 1st, 2018), but sulphur dioxide emissions from TPPs included in the national reduction plans (NERP) of North Macedonia, BiH, Serbia, and territory of Kosovo still exceeded five to six-fold than allowed. Of all the domain components, if air quality continues to be avoided as a problem and continues to worsen, it has the most significant potential to provoke mass protests and undermine peace. Third, tensions surrounding the war in Ukraine keep the issue of energy supply sustainability high on the political agenda. Energy supply sufficiency was raised as a profound national and geostrategic priority in EU27 and Western Balkans in 2023. The WB governments placed energy security, which is just a fraction of the energy performance indicator, considerably higher than any other component of the entire environmental sustainability domain on their political agenda in 2023. Energy system performance ranked high in Albania and Croatia due to diversified industries and a higher percentage of renewables in total final energy consumption, and moderate in Montenegro.

Dr. Dagmar Simon
Evaconsult and Social Science Center Berlin, Germany

EVERYONE IS EQUAL IN THE LOTTERY DRUM

ABSTRACT

One of the most striking changes in research funding systems in recent years is the greatly increased importance of third-party funding for research financing. In the context of New Public Management, the acquisition of third-party funding has become a significant factor in the governance of science, for example as an incentive in the performance-oriented allocation of funds by universities and as an important indicator for the assessment of research performance in evaluations. This development has shaped almost all European science systems in recent years. In institutional evaluations, performance dimensions are now recorded using a broad range of indicators. There is often a hierarchy with articles in refereed journals in first place and thirdparty funding in second place.

Added to this is the overloading of peer review, which has now become an internationally observable problem. New topics such as transfer activities, infrastructures and governance and, last but not least, complex and time-consuming review processes such as those in the Excellence Initiative are being addressed.

Despite all the pressures, peer review continues to be the foundation of scientific assessment. In peer review, the quality expectations and quality standards of the scientific communities are asserted, which, through their evaluations, award the decisive currency in science, namely reputation. As recognized and without alternative as the procedure is, studies have nevertheless long since drawn attention to structural problems and pointed to a lack of agreement between reviewers (reliability) and validity as well as (among other things, genderspecific) bias and the Matthew effect. In addition, in recent years in particular, peer review, which is largely discipline-oriented, has increasingly had to deal with inter- and transdisciplinary reviews of journal manuscripts and grant applications and develop criteria and (new) procedures for this. Another point of criticism of peer review that is of particular interest here is the structural conservatism of the reviewers, which makes it particularly difficult for new research ideas that cannot refer to the state of the art in the relevant research field or are otherwise secured to be reviewed and to successfully acquire third-party funding.

The lecture will ask how the Volkswagen Foundation has responded to these problems of review processes, which are now widely discussed in academia, with its “Experiment!” funding initiative, and how the funding recipients assess the initiative.

The aim of “Experiment!” was to support research projects that deal with particularly risky and original research questions. With the exception of the restriction that applications should come from the natural sciences, engineering and life sciences, no content

requirements or thematic priorities were set. Applicants were expected to already hold a doctorate and be employed at either a university or a non-university research institution in Germany. The people finally selected were provided with funding of 120,000 euros for a maximum period of 18 months.

The funding was open-ended; unexpected findings and even failure of the project were accepted as outcomes. From the outset, all calls for proposals met with a very high level of interest. In 2017, the selection process was fundamentally changed: In addition to selection by a jury, roughly the same number of applicants were now selected by lot. Since then, the number of funded projects has almost doubled, but the number of applications has also continued to rise:

In addition to focusing on new and risky research ideas in terms of content, the Volkswagen Foundation has also broken new ground in the design of the application and selection process. Applicants were expected to submit a short, maximum three-page text explaining their idea. This idea had to be completely new, so applicants were not expected to have done any preliminary work or even published work by other scientists. In addition, the three-page outline should be completely anonymized.

The jury was made up of eight to ten internationally recruited scientists. They were not recruited on the basis of their professional proximity to individual fields or in their capacity as specialists for specific issues, but as generalists for a broad spectrum of new and promising research ideas.

As part of the accompanying research, online surveys were conducted from the 2013 - 2019 approval rounds. In addition, around 40 guided interviews were conducted with selected grantees between 2019 and 2020, which were transcribed and analyzed using qualitative content analysis methods.

In view of the continuing tension between the breakthrough of new ideas and paradigms in science on the one hand and the structural conservatism of the peer review process on the other, the funding initiative offers the opportunity to actually try out something new that other research funding organizations or other funding programmes would not give a chance. In addition to the anonymized selection process, the partially randomized procedure in particular increases the chance of getting unconventional and risky projects off the ground. In addition, the majority of respondents had a positive assessment of lottery procedures, not least in view of the weaknesses of peer review procedures. A lottery procedure is particularly suitable for scientists in an early career phase who are not yet well established in the scientific community.

Another advantage for the assessment of risky research is seen in the composition of the jury: a small, international and interdisciplinary group that is responsible for three major scientific fields - life sciences, natural sciences and technical sciences. Against this background, the jury must focus on overarching questions such as whether the research project is both risky and feasible.

Among other things, research on peer review addresses the problem of bias in the assessment of research proposals: it can be assumed that the anonymization of applicants contributes to greater diversity. Specifically, another element of "Experiment!", the lottery procedure, shows that effects on diversity with regard to age, career stage and gender ratio can be recognized and can thus counteract a possible bias in the review process. The cohort whose projects were selected exclusively by the jury in the first four years of the "Experiment!" funding line was compared with the funding cohort from 2017 onwards, in which the partially randomized procedure was introduced. A combined evaluation of the age and gender of the grantees is revealing. It shows that the proportion of women has increased since 2017, in particular the participation of established female scientists (over 50 years old) and young female postdocs (under 35 years old) has increased. Overall, the proportion of funded persons under the age of 39 has increased in the partially randomized procedure. In terms of career stages, the proportion of researchers in early career phases (postdocs and junior professorships) has increased slightly. Lottery procedures therefore have a certain potential to counteract an age and gender bias.

"Experiment!" is catching on: Overall, the funding organizations have become bolder: For example, after a pilot phase, the Swiss SNF has now offered the option of a lottery procedure for all funding programmes, the British Academy is using partially randomized selection for smaller funding projects in the social sciences and humanities, and the Danish Novo Nodisk Foundation is also experimenting with the anonymization of applications in some funding lines. In general, a certain openness towards new funding formats, which may also contain experimental elements, can be observed in the European research funding systems. This is mainly due to the fact that a variety of formats and orientations of research funding is conducive to fairer participation opportunities for applicants, as this can compensate for deficits in one funding format with others.

Erik Canton
European Commission, DG for Research and Innovation, Belgium

A PORTFOLIO APPROACH TO RESEARCH FUNDING

ABSTRACT

Research funding decisions have long-term implications as they shape the research and innovation landscape of the future. There is however remarkably little attention in the policy debate to the question how to make such decisions, a question becoming even more pertinent in an increasingly complex political environment with multiple (and sometimes unstable and/or conflicting) objectives for research and innovation (R&I4) policy.

This study argues that portfolio theory can provide a powerful tool to make research funding decisions in a complex environment. The main idea is simply that instead of deciding on the basis of the contributions of individual proposals, one can allocate funding by looking at the attributes of the combined set of proposals ultimately selected. The portfolio approach in this context simply refers to the maximisation of an objective function under a set of constraints, using linear or non-linear programming and a heuristic procedure to find maximum values. The quality of research projects is an important attribute in this procedure, but it is not the determining factor.

The main advantage of such a portfolio approach is that it can also take account of interdependencies between research proposals (substitutability/complementarity and risk diversification) and of holistic attributes referring to other objectives (for example related to inclusiveness and diversity considerations or whether the proposal is addressing a topic relevant for progress towards the Sustainable Development Goals). The approach put forward in this article can be seen as a management information tool, allowing research funding decisions to be taken from a systematic overview of alternative outcomes in a fully transparent process.

The prevailing method is to make funding decisions based on the rankings of the received research proposals (cf. Banal-Estañol et al., 2019). The advantage of this method is that it is merit-based, funding goes to the proposals with the highest scores from the evaluators. Discretionary adjustments can be made to cater possible other objectives. But this is typically done in an ad-hoc, mostly qualitative and nonsystematic manner, with subjectivity coming into play. Such an easy to apply funding method may not necessarily lead to the best possible outcome.

A possible explanation for the widespread use of merit-based funding is that the decisionmakers follow conventions and codes of conduct typically motivated by fairness considerations in the sense that the projects with the highest scores are funded. Deviations from such a merit-based approach could be seen as going against conventions, possibly

with negative consequences (such as criticism from peers). Results from a survey among policy practitioners from the European Commission provide informal evidence for the presence of such conventions in research funding decisions. Practitioners in the field of research and innovation policy who are more exposed to such conventions showed a stronger preference for the merit-based approach than policy practitioners active in other fields, who are likely less influenced by the research and innovation community's conventions and codes of conduct.

This article builds upon various strands in the literature, which I will concisely review. In particular, I will briefly summarise the evolving rationale for R&I policy, which also helps to explain why funding decisions have become more complex. Then I will review some main references of portfolio thinking in research funding decisions, and make a link with institutional economics to explain the discrepancy between theory and practice, where merit-based funding is the prevailing method.

Rationale for R&I policy: The economic rationale for R&I policy has developed along three types of arguments. Firstly, public intervention in the field of research and innovation is needed to correct for market failures. An oftenmentioned market failure is knowledge spillovers. Knowledge has public good properties, in the sense that it is non-rival and only partly excludable (Romer, 1990). This implies that private investors in research projects generating new knowledge can only appropriate part of the returns, which will lead to underinvestment from a social perspective (see for example Jones and Summers (2020), reporting that the social returns to innovation are very large). Public interventions, for example in the form of grants or intellectual property protection, will reduce the wedge between private and social returns, helping to internalise knowledge spillovers. The discrepancy between private and social returns can also be reduced through the presence of a well-functioning secondary market (cf. Arqué-Castells and Spulber, 2022). Other market failures identified in the literature refer to information asymmetries causing access to finance difficulties (in particular for SMEs), indivisibilities of research projects and other nonconvexities, and difficulties to wash out risk in sets of research projects because of a-typical risk patterns of individual projects. All these market failures can lead to lower private investment in R&D than socially optimal, justifying some form of public intervention to try and restore socially optimal investment levels. The notion of market failures goes back to the early days of neoclassical theory, and its application in the field of research and innovation was spurred through the development of endogenous growth theory in the late 1980s. For more detailed discussions, see for example Hall (2002), Hall and Lerner (2009), and Mazzucato and Semieniuk (2017).

Secondly, in the 1980s a literature emerged on system failures, where interdependencies between various parts of the innovation system could lead to underperformance in case of bottlenecks. The main message here is to not only consider R&I activity in isolation, but to look at the wider system in which such activity is embedded. A popular example of system failure is the lack of linkages between science and industry (culminating in the "European paradox", where it is claimed that Europe shows strong scientific performance but lags behind in the ability of converting this into wealth-generating innovations, cf. Dosi et al. (2006)). Another example would be when the effectiveness of research subsidies in terms of increased research activity is limited due to a mostly inelastic supply of researchers (at least in the short run). To mitigate upward wage pressures, efforts to increase the supply of researchers (for example by encouraging students to enrol in a science or engineering programme or targeted immigration policies such as the EU's Blue Card system) should then be taken before stepping up financial support to the private sector to engage in R&I activity. As a final example, system failure can refer to the functioning of public administrations and the design of public policies ("government failure"), where gains can be reaped by systematically using robust evidence in favour of or against a certain type of intervention (evidence-informed policymaking), and by ensuring synergies with other intervention areas which are also part of the wider innovation system. For further discussions, see for example Lundvall (1992). The extent to which proposals are addressing such system failures can be taken on board in research funding decisions.

Thirdly, in the most recent R&I literature there is a call to make R&I policy more transformative. A pioneering paper is Schot and Steinmueller (2018) who argue that "to meet the ambitious challenges expressed for example in the SDGs, we need a new framing for innovation policy. This is what we call Framing 3 aimed at transformative change. This raises the question – what needs to be transformed? Based on the research in sustainability transitions studies we argue that transformation of socio-technical systems is needed in energy, mobility, food, water, healthcare, communication, backbone systems of modern societies" (page 1562). The definition of transformative research and innovation policy has further broadened in debates on a renewed growth model, with a stronger emphasis on inclusiveness, sustainability, resilience, open strategic autonomy and preparedness. Transformative research and innovation policy thus refers to a broad policy agenda with the ultimate objective to address the major societal challenges mankind is facing. Discussions on how such transformative policy should look like are ongoing, with a recurrent role for directionality in research programmes to channel more funding to projects highly relevant for tackling such major societal challenges. Lepori et al. (2023) discuss in greater detail the changing rationales for funding towards grand challenges.

A related strand in the literature is about the measurement of these additional criteria. For example, Bozeman and Youtie (2017) discuss the evaluation of non-economic social outcomes of research, pointing (inter alia) at the potential of bibliometrics to track social value. There is also a literature on how R&I policy shapes research content (cf. Gläser and Laudel (2016) for a review). For example, Røttingen et al. (2013) and Yegros-Yegros et al. (2020) show that research investments in the health sector are concentrated in diseases or topics driven by economic and/or political factors, thereby illustrating how incentives and the institutional environment matter for the directionality of investments in R&I.

With the set of policy objectives expanding over time, research funding decisions are becoming increasingly complex. Sarewitz and Pielke (2007) investigate how to reconcile science supply and societal demand, offering strong arguments on why it is important to have a portfolio level assessment.

Portfolio approach: The portfolio approach is mainstream in the world of finance and investment (since the work by Markowitz, 1952). In the finance community it is a statistical method to improve the properties of an investment portfolio. Investments are diversified in terms of risk categories and assets within each risk category. Exploiting the law of large numbers, total risk levels decrease when diversification increases within and across risk categories. The investment decision on a particular asset is therefore not made in isolation, but is contingent on the composition of the rest of the portfolio. The abundance of financial data on correlations, variances and co-variances of asset returns has spurred academic research and practical applications of the portfolio approach in the financial sector.

References to portfolio theory appear regularly in the R&I policy debate. For a variety of reasons the portfolio approach from the finance literature cannot be translated one-to-one to the world of R&I.5 The finance literature using portfolio theory is largely based on quantitative modelling and econometric analysis, whereas the inclusion of portfolio thinking in research funding decisions is typically done in a more qualitative fashion. There are however several studies that can help applying more rigorous quantitative methods to research funding decisions. Dorfleitner et al. (2012) stay close to the original approach and introduce a social dimension in the Markowitz portfolio model, where assets both generate a financial return and a social return.6 They look at stochastic social returns, and at a simplified version with deterministic returns. The advantage of the former is that the portfolio performance can be improved by exploiting the covariance structure between various assets. Linton et al. (2002) use Data Envelopment Analysis to build a portfolio of R&D projects. Hall et al. (1992) and Chien (2002) use linear algebra techniques to select projects in a deterministic environment, and showcase the flexibility of this approach when there are multiple objectives.

Conventions and institutional economics: The portfolio approach is not commonly used in research funding, and the prevailing method is to fund projects on the basis of their individual quality score from the evaluation process (cf. Banal-Estañol et al., 2019). This article explores the role of conventions as a potential explanation for this discrepancy and is thereby related to the literature on institutional economics. According to North (1994, page 360), "Institutions are the humanly devised constraints that structure human interaction. They are made up of formal constraints (e.g., rules, laws, constitutions), informal constraints (e.g., norms of behavior, conventions, self-imposed codes of conduct), and their enforcement characteristics. Together they define the incentive structure of societies and specifically economies." The convention to use meritbased funding can be seen as an informal constraint with an impact on outcomes. This paper is thereby also (but somewhat more remotely) related to the literature on the role of informal constraints for the decision process, for example on the effectiveness of codes of conduct (as an alternative to regulation), the influence of social norms on behaviour (cf. Bicchieri et al., 2023), and the economics of convention (cf. Diaz-Bone, 2018).

The motivation behind this study is to investigate whether funding decisions can be improved by moving away from the convention of deciding on individual proposals towards an approach where the total impact of the call for proposals is maximised.
A large advantage of the portfolio approach is that decisionmakers can be canvassed in the design stage of the funding procedure, and the procedure can be tailor-made to accommodate specific requirements and desirable features (e.g. as regards interdependence of proposals and synergistic attributes in a multicriteria setting). One would have to give up the ex ante commitment to fairness central to the meritbased approach, but in return there are important benefits in the sense that management decisions can be made in a more transparent and systematic way, taking on board some of the important complexities R&I policymakers see themselves confronted with these days.
Even for the basic model, in the absence of multiple objectives or other factors making prioritisation more complicated, portfolio thinking can improve research funding decisions. It can be proven mathematically that the outcomes from the portfolio approach are always at least equally impactful than the ones from the merit-based approach. The intuition is that one can obtain merit-based outcomes from an amended version of the portfolio model with constraints on funding order, and these constraints can be slack (in which case the outcomes are equally attractive) or binding (in which case the menu of proposals selected from the portfolio model outperforms the menu of proposals from the merit-based approach). The outcomes from the portfolio model in terms of total impact

are therefore always at least equally good as the outcomes from merit-based funding. Moreover, with the ability of the portfolio approach to consider complex facets such as substitutability/complementarity between projects, diversity, and multiple selection criteria, it becomes even more apparent that the portfolio model would enable better informed funding decisions.

Nino Gogadze, Lasha Macharashvili
University for Continuing Education Krems, Austria
Tampere University, Finland

TOWARDS RESPONSIBLE RESEARCH ASSESSMENT: INSIGHTS FROM POST-SOVIET COUNTRIES

ABSTRACT

In recent years, Responsible Research Assessment has gained growing importance within the European Research Area and has emerged as a focal science policy objective. In 2021, the European Commission (EC) released a scoping report and proposed a coalition approach for research assessment reform (European Commission, 2021), which set the groundwork for the establishment of the Agreement (2022) and Coalition for Advancing Research Assessment (CoARA) in Europe (CoARA web site, 2024). This development is also relevant for Georgia, which was granted candidate status for EU membership in December 2023 and is an associated country of Horizon Europe (European Commission, 2024). While studies confirm the need to reform the existing research evaluation system in Georgia (State Audit Office of Georgia, 2014; Tabatadze & Chachkhiani, 2022; Macharashvili & Gogadze, 2023; Tsotniashvili, 2023), as of January 2024, only three organizations from the country have joined the CoARA agreement: two private higher education institutions (out of a total of 56) and the main research funding body —the Shota Rustaveli National Science Foundation of Georgia (CoARA Web site, 2024).

Nowadays, research evaluation has become a challenge in Georgia, where, due to the Soviet legacy, university-based research is a relatively new phenomenon, as in the Soviet Union, the research was predominantly carried out by the specialized research institutes operated under the Academies of Sciences, resulting in a distinct separation between teaching and research activities, with higher education institutions playing a minor role in the production of new knowledge (Chakhaia & Bregvadze, 2018; Chankseliani et al., 2021; Chankseliani, 2022; Lovakov et al., 2022; Tabatadze & Chachkhiani, 2022; Macharashvili & Gogadze, 2023; Tsotniashvili, 2023).

Despite efforts to align with European standards, the research assessment system remains fragmented, inefficient and ineffective (State Audit Office of Georgia, 2014; Tabatadze & Chachkhiani, 2022; Macharashvili & Gogadze, 2023). The need for comprehensive reform in research assessment practices and support for early career researchers in Georgia is increasingly evident (Dzotsenidze, 2022; Tabatadze & Chachkhiani, 2022; Macharashvili & Gogadze, 2023). Given that the CoARA Agreement is a recent development, this study is the first attempt to explore its implementation in Georgia.

The purpose of this study was twofold: first, it aimed to explore the implementation of the CoARA agreement in Georgia, and second, to examine the challenges faced by Georgian

signatories to the CoARA agreement. To achieve the research objectives a case study design was employed. The study employed two different data collection methods: desk research for secondary data and semi-structured interviews for primary data collection. The documents analyzed during the desk research included: legislation, strategic and policy documents, institutional policies of the National Center for Educational Quality Enhancement and the Shota Rustaveli National Science Foundation of Georgia. The second phase of the research involved interviews with representatives from the CoARA agreement signatory organizations and other relevant stakeholders involved in research assessment. In total, 5 interviews were conducted, Table 1 provides a summary of the main characteristics of the participants:

TABLE 1. CHARACTERISTICS OF THE STUDY PARTICIPANTS

Country	Organization	Signatories of CoARA Agreement	Management Level
Georgia	Petre Shotadze Tbilisi Medical Academy	YES	Top
	David Tvildiani Medical University	YES	Middle
	National Center for Education Quality Enhancement (NCEQE)	NO	Top
	Ministry of Education and Science of Georgia	NO	Middle
	Shota Rustaveli National Science Foundation (SRNSFG)	YES	Middle

*Following the document analysis, interviews were conducted between February 28 and April 24, 2024.

According to the study findings, the state of research assessment in Georgia reflects a mixed landscape of progress and challenges. While discussions around research assessment are prevalent, Georgia is in the nascent stages of European responsible research assessment reform. Furthermore, Georgia's research evaluation system incorporates elements from both weak and strong research evaluation ideal models, as defined by Whitley (2007). While lacking a single, unified system for research assessment, Georgia utilizes multiple entities such as the Georgian National Academy of Sciences, the National Center for Educational Quality Enhancement and the Shota Rustaveli National Science Foundation of Georgia for research evaluation from different perspectives. Moreover, the research indicates a lack of awareness regarding responsible research assessment reform, which extends even to signatory organizations. For example, the SRNSFG has undergone three leadership changes in the past two years, resulting in insufficient dissemination of information among its staff regarding the foundation's commitment to the CoARA agreement and its responsibilities. Additionally, no information regarding the signing of the agreement can be found on the official websites of Georgian organizations, highlighting a broader communication gap in disseminating this information. Considering all of the above, when applying the typology of responses to RRA initiatives developed by Curry et al. (2020), such as "Cosmetic appropriation," "Calibrating the machine," "Can openers," "Advocacy Coalitions," and "Institutional Culture Change," the case of Georgia emerges unique and does not fit into any of them.

While various policy documents express the government's willingness to improve research assessment policies through the introduction of external quality assurance measures for research institutions and the implementation of performance-based funding, in these documents, research assessment emerges within the New Public Management (NPM) discourse, emphasizing the importance of outcomeoriented results, increased efficiency, accountability and performance. However, the CoARA agreement goes beyond the narrow focus of NPM by placing significant emphasis on open science, collaboration, diversity, citizen science, equity, inclusion, societal relevance and impact in research assessment (CoARA web site, 2024), Therefore, both the signatories of the CoARA and Georgian policymakers should embrace this broader perspective in their approach to research assessment.

Furthermore, the results indicate that the challenges facing the implementation of the CoARA agreement in Georgia are multifaceted. These include existing national research assessment policies that diverge from the CoARA principles, incomplete integration of scientific research staff within universities, resistance from senior academics, difficulties in balancing qualitative and quantitative assessment indicators, and skepticism towards the agreement's applicability, as it may introduce further ambiguity, with the agreement being seen as more relevant for European countries and the USA.

Given the recent development of the CoARA agreement, the study enriches the existing data and sets a foundation for further studies in this area. Additionally, at a broader level, the study contributes to the ongoing global discourse on responsible research assessment and provides insights regarding the implementation of the CoARA agreement in non-EU and post-Soviet contexts. By exploring its implementation in non-EU contexts, it offers insights into the challenges and opportunities faced by countries outside the European Union, thereby benefiting the global research community.

REFERENCES:

- Chankseliani, M., Lovakov, A., & Pislyakov, V. (2021). A big picture: Bibliometric study of academic publications from post-Soviet countries. Scientometrics, 126(10), 8701–8730. https://doi.org/10.1007/s11192-021-04124-5

- Chankseliani, M., (2022). What Happened to the Soviet University? Oxford University Press.
- Chankseliani, M., Fedyukin, I., & Frumin, I. (Eds.). (2022b). Building Research Capacity at Universities: Insights from Post-Soviet Countries. Springer International Publishing. https://doi.org/10.1007/978-3-031-12141-8
- CoARA - Coalition for Advancing Research Assessment. CoARA. Retrieved May 27, 2024, from https://coara.eu/
- Curry, S., de Rijcke, S., Hatch, A., Pillay, D. (Gansen), van der Weijden, I., & Wilsdon, J. (2022). The changing role of funders in responsible research assessment: Progress, obstacles and the way ahead (RoRI Working Paper No.3) [Report]. Research on Research Institute. https://doi.org/10.6084/m9.figshare.13227914.v2
- Dzotsenidze, N. (2022). Barriers to completion of PhD studies in Georgia. [Policy Paper]. PMC Research Center.
- European Commission. Directorate-General for Research and Innovation. (2021). Towards a reform of the research assessment system: Scoping report. Publications Office of the European Union. https://data.europa.eu/doi/10.2777/707440
- European Commission. Web site (2024, May 15). Retrived from: https://neighbourhood-enlargement.ec.europa.eu/european-neighbourhoodpolicy/countries-region/georgia_en
- Government of Georgia. Government Programme 2021-2024: Action Plan of the Development Strategy of Georgia - Vision 2030, web site https://www.gov.ge/
- Government of Georgia. Government Programme 2021-2024: Toward Building a European State, web site https://www.gov.ge/
- Government of Georgia. Development Strategy of Georgia - Vision 2030, web site https://www.gov.ge/
- Legislative Herald of Georgia. Web site http://matsne.gov.ge/
- Lovakov, A., Chankseliani, M., & Panova, A. (2022). Universities vs. research institutes? Overcoming the Soviet legacy of higher education and research. Scientometrics, 127(11), 6293–6313. https://doi.org/10.1007/s11192-022-04527-y
- Macharashvili, L., & Gogadze, N. (2023). The many faces of research assessment: Case of independent Georgia. Journal of Research and Innovation in Higher Education, 4(1), 1–27. Retrieved from https://rihe-journal.com/index.php/rihe/article/view/31
- Ministry of Education and Science of Georgia (MoES). Web site https://mes.gov.ge/
- Ministry of Education and Science of Georgia (MoES). Unified National Strategy of Education and Science of Georgia for 2022-2030. Web site https://mes.gov.ge/

- National Center for Educational Quality Enhancement (NCEQE). Web site https://eqe.ge/
- Shota Rustaveli National Science Foundation of Georgia (SRNSFG). Web site https://rustaveli.org.ge/eng
- State Audit Office of Georgia. (2014). მეცნიერების მართვის ღონისძიებების ეფექტიანობის აუდიტის ანგარიში [The efficiency report of the science management system]. Tbilisi. Retrieved from https://sao.ge/files/auditi/efeqtianobis-angarishi/2014/mecniereba.pdf
- Tabatadze, S., & Chachkhiani, K. (2022b). University-Based Research and Development in Georgia. In M. Chankseliani, I. Fedyukin, & I. Frumin (Eds.), Building Research Capacity at Universities: Insights from Post-Soviet Countries (pp. 201–221). Springer International Publishing. https://doi.org/10.1007/978-3-031-12141-8_11
- Tsotniashvili, K. (2023). Transformation of Academic Identities in Georgia: Diffractive Narratives In-Between Defuturing and Refuturing, Dissertation/Thesis. Arizona State University.
- Whitley, R. (2007). Changing Governance of the Public Sciences. In R. Whitley & J. Gl ‰ ser (Eds.), The Changing Governance of the Sciences: The Advent of Research Evaluation Systems (pp. 3– 27). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6746-4_1

Anke Reinhardt
Deutsche Forschungsgemeinschaft (DFG), German Research Foundation, Germany

DEVELOPMENT OF THE EVALUATION ACTIVITIES OF THE DEUTSCHE FORSCHUNGSGEMEINSCHAFT (DFG) - EXAMPLES OF RECENT PROJECTS

ABSTRACT

The purpose of this presentation is to explain the DFG's evaluation framework and recent developments. In particular, I will give examples of recent evaluation activities.

What are the DFG's evaluation activities?

The DFG's evaluation activities include the collection of descriptive statistics and information on the implementation and impact of DFG funding. The DFG's evaluation activities are concerned with the "reality" of funded projects (as opposed to the ideas and intentions formulated by researchers in project proposals) and the effects (outcomes and impacts) of funding. Like many funding agencies (Safl 2010, Helgesen & Reinhardt 2012), the German Research Foundation (DFG) has implemented a specific evaluation framework tailored to its needs (Reinhardt 2013). As part of its evaluation activities, the DFG conducts surveys (recurring or one-off) and commissions studies (usually entire studies or occasionally individual modules) from external contractors (Reinhardt & Geyer 2016).

Evaluation activities contribute to the discussion of funding programmes and policies by showing how the research landscape is developing and what is working. They thus improve the information base available to decision-makers and decision-making processes within the DFG. As evaluation studies are made publicly available, they also communicate the scientific and structural contribution of research funding to policymakers and the interested public.

What are the cornerstones of the DFG's evaluation activities?

There are several characteristics and cornerstones of the DFG's evaluation activities.

The DFG conducts evaluations primarily to learn about its own programmes. This distinguishes the DFG from other funding organisations, which use evaluations primarily to report to their sponsors on the effectiveness of their research funding. The DFG has decided not to carry out repeated evaluations of its funding programmes. Instead, each individual evaluation focuses on a specific, predefined topic. Evaluation studies are only commissioned if they are expected to provide insights that will usefully contribute to decision-making. Compared to other organisations, the DFG has a very quantitative approach. This is partly due to the practice that evaluation studies only form the basis for an assessment of a programme, but that the assessment in the form of conclusions on the report is made by evaluation commissions and the decisions are taken by the statutory bodies of the DFG.

This also leads to the use of a variety of methods, including bibliometrics, surveys, interviews, case studies and document analysis.

Developments within the DFG's evaluation system: The DFG's evaluation system has evolved over the past ten years. This can be seen both as a deliberate change and as a reaction to external and internal requirements or developments. I would like to highlight three developments and how they are reflected in the DFG's evaluation activities.

1. Increasing use of scientific evidence: As the complexity of the research and funding landscape increases, so does the need to understand the functional logic of research funding. As a rule, evaluation itself does not claim to produce generalisable findings. Instead, individual evaluation projects have a specific question and attempt to find answers to it. Typically, evaluation studies are carried out by professional evaluation agencies, i.e. service providers. Nevertheless, evaluation and research are converging.

Researchers, driven by their scientific curiosity, pose research questions of interest to funding agencies, and funding agencies hope to benefit from the methodological skills, outside perspective and originality of researchers. Driven by the idea of evidence-based decision making, decision-makers increasingly want to base their judgements on a sound information base that can only be generated through the application of scientific methodology.

In the case of the study on the Koselleck programme, the DFG decided to commission a study from scientists. It issued a closed call to a range of researchers from the science of science community, offering access to data on funding proposals and a small amount of money to cover expenses. Two teams were selected on the basis of the originality and relevance of their proposals. The expectation of the research team is that they will submit a working paper to the DFG, but the main goal is that they will then submit this or more articles to scientific journals.

The DFG hopes to gain insights to questions that were not imagined by itself. by In particular, this tendering process was new ground.

2. Improving the database: With the advent of more databases, new software and new technologies, the database available for evaluation studies is improving dramatically. However, this still requires a good understanding of the quality of the data content and investment in data quality and analysis. I will give two examples of how the DFG is expanding its database. Bibliometric studies on the outcomes of research funding often use the Funding Acknowledgement field in publication databases (e.g. Web of Science, Scopus). In the past, the reliability and validity of these analyses were unknown and seemed to have some weaknesses. Therefore, the DFG commissioned a study to investigate whether and how well DFG project participants comply with the DFG's rules on naming the DFG and including the specific project identifier in funding acknowledgements.

The analyses are based on information from more than 50,000 publications cited as project results in DFG final reports and the relevant texts of the funding acknowledgements as contained in the publication database "Web of Science". The study also developed a search strategy that can be used by the bibliometric community.

In addition, the DFG has been investing for several years in creating the basis for text mining activities. While most of the DFG's evaluative and statistical products and services are based on structured data from the electronic processing of proposals or from surveys, the DFG also stores documents (proposals, reviews, final reports, CVs) in an internal document management system. The DFG has now started to extract the textual information, convert it into a database, process and clean it, analyse it and create information products.

3. Further professionalisation The DFG's evaluation team has grown considerably over the past few years, and its expertise has also broadened. In addition, new internal DFG organisational units provide services that play a role in the process of commissioning evaluation studies, such as a team for data protection or a team for procurement law.

The evaluation team also wants to benefit from the knowledge of other actors in the field and is therefore actively involved in various national and international networks, such as the network of evaluation officers in funding agencies in Germany (Heidler et al. 2016), the GRAIL working group of the Research on Research Institute, Science Europe, and the G7 Working Group on Research Assessment.

In summary, the DFG's evaluation activities have developed into a multifaceted bouquet. Classic programme evaluation, ranging from commissioned programme evaluations to external evaluation studies, still exists, but is the exception rather than the rule. Instead, the mix includes scientific studies, methodological assessments and a diverse database. This, of course, requires new skills and a continuous process of training and refinement.

REFERENCES:

- Heidler, R., Reinhardt, A., & Safl, U. (2016). „Evaluation der Fˆrderung von Wissenschaft und Forschung–R¸ckblick, Einblick, Ausblick “am 16. Juni 2016 in Bonn. Zeitschrift f¸r Evaluation, 15(2), 306-313.
- Helgesen, G., & Reinhardt, A. (2012). A Guide to Evaluation Activities in Funding Agencies. ESF Member Organisation Forum on Evaluation of Publicly Funded Research.
- Reinhardt, A. (2013). Das Evaluationsvorgehen Der Deutschen Forschungsgemeinschaft. Zeitschrift f¸r Evaluation, 12(2), 305-314.

- Reinhardt, A. & Geyer, A. (2016). Everyone’s happy? The case for expectation management in commissioned evaluation projects. ftEval Journal for Research and Technology Evaluation. Issue 42.
- Safl, U. (2010). Evaluationen–Instrumente der Qualit ‰ tssicherung der VolkswagenStiftung. Zeitschrift f¸r Evaluation, 9(1), 143-152.

New Methods, Tools and Implications for Evaluation in the Digital Era

Muhammad Asaduzzaman
University of Oslo, Norway

DISTRICT HEALTH INFORMATION SOFTWARE (DHIS2) FOR ONE HEALTHANTIMICROBIAL RESISTANCE SURVEILLANCE: QUALITATIVE EVALUATION OFTHE FEASIBILITY AND ACCEPTABILITY

ABSTRACT

Antimicrobial resistance (AMR) is currently one of the major global One Health (OH) challenges that acknowledges the interdependence between human health, animal health, and our shared ecosystem for its containment. The irrational use of antibiotics and the presence of multidrug-resistant (MDR) pathogens in multidisciplinary OH settings such as humans, food animals, including aquaculture, wildlife, and the environment are well documented. Therefore, the success of containment of the local, national, and global spread of AMR depends largely on a technically sound, timely, and integrated surveillance system. Only a comprehensive surveillance system can afford required data for all aspects of AMR, such as the effectiveness of antibiotic therapies, stewardship programs, or programmatic interventions in various settings. Unfortunately, existing local or global AMR surveillance systems lack an integrated OH approach, which means synchronised human, animal, and environmental resistance data in the same surveillance network are missing.

To achieve an effective interdisciplinary approach, the surveillance platform must be capable of flexible and uncomplicated AMR data archiving, analysing, and sharing. The digital health information system (HIS) has enhanced the quality and cost-effectiveness of disease or service reporting through health system surveillance to a greater extent. A digital HIS-supported AMR surveillance system can be more holistic in nature, having an effective role in identification of better treatment modalities, monitoring of AMR trends, policy guidance, and real-time resource allocation. HIS usually deals with either health data at the personal level, such as electronic medical records, or health system data at the population level for decision-making, such as district health information software, version 2 (DHIS 2). DHIS2 was chosen as the surveillance platform to be investigated in our study for multiple reasons. Firstly, it is the leading global health information system that fulfils the criteria of a digital global public good (freely available) and is already utilised by more than 60 countries for health system management.

LMICs are in need of eminent health information systems but do not have the required technical capacities, logistics, and tools. This in turn may result in flawed data quality, inappropriate policymaking, and resource allocation for major health issues, including AMR. AsDHIS2 allows to report health indicators in real time from local health facilities with data deposition in the central server, more integration of DHIS2 in the HMIS of developing countries is much required, especially in South Asia and Africa. Therefore, the

continuous reporting by remote health facilities through DHIS2 is always helpful to measure the program impacts and required interventions. However, the previous experiences of implementing DHIS2 in several countries identified multiple challenges, mainly technical problems and stakeholders' acceptance and preparedness. As there is no feasibility data regarding this system exists in Ethiopia, such a study will aid in the understanding of information and on job training needs for the system. We also believe that such studies will enhance the awareness and knowledge among the relevant scientific community regarding OH-AMR and its integrated surveillance system. Although many African countries (e.g., Tanzania, Zambia, Ethiopia, Kenya, Ghana, Liberia, South Africa, Malawi, South Sudan, Cameroon, and Nigeria) are currently using DHIS2 in their health system reporting, the infrastructure, resource allocation, and trained personnel capacity for digital platforms vary a lot, which will also impact the outcome of the OH-AMR data. Therefore, we aimed to conduct an exploratory qualitative study for understanding the feasibility, acceptability, and perceptions of the DHIS2 among local OH-AMR stakeholders in Jimma, Ethiopia.

This study was conducted in the Jimma Zone of the Oromia region in Southwest Ethiopia, which is subdivided into 21 administrative units called Woredas (districts). This zone has approximately 3.3 million inhabitants. Our study population was mainly from Jimma town, which is 346 km south-west of Addis Ababa, the capital of Ethiopia and the largest city in the region. A qualitative cross-sectional study design comprising key informant interviews (KIIs) was conducted from November 2020 to February 2021. Data obtained from KIIs helped to understand the key informants perception and acceptability of an integrated digital AMR surveillance platform such as DHIS2 and the feasibility of its context-specific establishment. It also provided important insights into the challenges of this platform from some of the current users.

Semi-structured interviews were performed in English among key informants in Jimma who are OHstakeholders for local AMR surveillance, mainly from academic institutions, local health offices, and relevant project staffs (Table 1). The participants provided their response based on their experience and knowledge. Table 1: Key informants' distribution based on one health discipline

Professional background	*Number of respondents*	*Sampling procedure*
Physician/Public health professionals (P)	*8*	*Purposive*
Veterinarian (V)	*5*	*Purposive*
Environmental professionals (E)	*5*	*Purposive, snowball*
Administrative personnel in academia (A)	*8*	*Purposive, convenient*
Data managers (D)	*8*	*Purposive, snowball*
Laboratory experts (L)	*8*	*Purposive*
Total study participants	*N=42*	

Following verbatim transcription of the audio recordings of KIIs, NVivo 12 software (QSR International, Melbourne, Australia) was used for line-by-line coding and analysis of the anonymised data. Thematic analysis was conducted, which included the understanding of the pattern of data, iterative coding to identify the major themes, and description of the themes by illustrative quotes.

The identified themes from KIIs helped to ascertain the 'programmatic feasibility', 'acceptability', and 'perception of DHIS2. The major themes and the obtained information are described below.

Theme 1: Knowledge about Digital Surveillance and DHIS

Before talking about DHIS2, the key informants were asked about their overall digital surveillance system knowledge. Most of the participants were familiar with the term and concept of digital surveillance systems, with limited or no prior experience of using them. However, fewer key informants were aware of the DHIS2 platform. 'Yes, I know EMR only—an electronic medical record; it was made, and I was part of it too. It is to give all reports to patients electronically. But I don't know about DHIS2. -(P1)'I am not aware of any AMR surveillance in our institute; also have no idea on how to use DHIS2.' –(A2) Considering the OH perspective, veterinarians and environmental professionals are important in the integrated system. In Jimma, academics and researchers from these two disciplines have limited knowledge on surveillance platforms.

Theme 2: Perceptions on digital platforms, including DHIS2

We received positive feedback from all key informants about the utilisation of digital platforms for any type of disease surveillance. The participants perceive the electronic reporting system, such as DHIS2, as a pathway of accessible, timely, and quality data flow. 'Yeah, that's out of question. That's (digital surveillance) very important because currently we tried the surveillance at the national level but it's not well coordinated.' –(A3) 'A digital platform like DHIS2, which will be user-friendly and easy to use with a very clear interface, will be important.' –(L2)

Theme 3: Suggested features in an OH-AMR surveillance platform and programmatic input

As several stakeholders in the local OH-AMR surveillance system had limited prior knowledge of DHIS2 as users, we received technical recommendations and coordination of the data synchronisation, monitoring, and evaluation from the current users mostly, especially from the data management professionals. Other key informants provided suggestions regarding possible data input about AMR in the system. Therefore, the recommendations are stated in three categories:

- Features recommended by the current DHIS2 users in Ethiopia;
- Recommendations for AMR data input in the customised DHIS2 platform;
- Recommended administrative and programmatic inputs from KIIs

Theme 4: Possible Implementation Challenges

Most of the key informants have identified some subjective barriers, such as lack of motivation and awareness at both the employee and institutional level, which need to be addressed before the DHIS2 program implementation. ‘The first thing is that we, as an institution, are not aware of the significance of this trait. If you are not aware and committed, you don’t allocate budget; you don’t commit yourself to the actions.’ –(A3)

Theme 5: Perceived role of the participants in the DHIS2 implementation phase

The key informants’ view on the implementation of the DHIS2 platform for OH-AMR surveillance is quite encouraging. The academic researchers have shown interest in DHIS2 mainly in three ways: 1) self-learning and utilisation in their projects; 2) collaboration with and dissemination of the platform to other research groups in the same organisation or different organisations at national or global level; and 3) utilisation as a teaching module for the students (in the form of learning by secondary data and using in postgraduate students' research). ‘My role will be supporting students and residents in usage of this platform, if it is friendly actually… and for the application in their practice. As I said, my role as a teacher too, so once I understand it, I pass it on to the students. Second, I will monitor the implementation of it as I am working as a clinician. And I will use it as apart in my research also, so we can develop better empirical management of the drugs.’ –(P4)

DHIS2 has the potential to synchronise all sorts of AMR data and to act as a standardised OH surveillance system. Based on the thematic analysis and study findings, the implementation of DHIS2 as an OH-AMR surveillance platform seems to be feasible both at local and national levels if the knowledge gaps and systematic barriers are addressed beforehand. An important reflection obtained from this study was the lack of coordination among the relevant stakeholders of AMR surveillance, both at the local and national level. A multidisciplinary orOH approach is not practiced in local AMR research or surveillance. Currently, no data sharing policy or mechanism exists at the study site, and the absence of animal health or environmental AMR surveillance in the national action plan may be one of the major reasons. The recent study on the AMR preparedness in SSA has also identified this gap and has recommended developing and implementing national action plans for OH-AMR surveillance. OH capacity building is particularly important for DHIS2 if it is designed to accommodate multidisciplinary AMR data. Taking the implementation barriers into consideration and solving them, DHIS2 can be the most user-friendly and acceptable interoperable platform in resource-poor settings. Piloting at local level and gradual scaling at national level with enhanced alignment among OH disciplines should be prioritised for the AMR containment.

Explanation of the relevance of the technical application of DHIS2 as an OH-AMR surveillance platform in policy-making or program evaluation context

Antimicrobial resistance (AMR) is a global One Health (OH) problem that is prevalent in human, animal, and environmental domains with transmission potential among the settings. Currently, there is no integrated surveillance platform to capture the OH-AMR data from all three domains. This solitary data capture practice is a major challenge to mitigate this global problem in holistic matters. District health information software, version 2 (DHIS2), is a globally recognised digital surveillance platform and is being used as health system surveillance and reporting in more than 100 countries. However, it has not been utilised as an AMR surveillance platform yet. The previous experiences of implementing DHIS2 as a health system reporting platform in several countries identified multiple challenges, mainly technical problems and stakeholders' acceptance and preparedness. As there is no feasibility data regarding this system as the OH-AMR platform exists, the qualitative feasibility study will aid in the understanding of information and on-the-job training needs for the system. We also believe that such studies will enhance awareness and knowledge among the relevant scientific community regarding OH-AMR and its integrated surveillance system. In addition, many African countries (e.g., Tanzania, Zambia, Ethiopia, Kenya, Ghana, Liberia, South Africa, Malawi, South Sudan, Cameroon, and Nigeria) are currently using DHIS2 in their health system reporting. The infrastructure, resource allocation, and trained personnel capacity for digital platforms vary a lot, which will also impact the outcome of the OH-AMR data. Before creating a new interface of DHIS2 for OH-AMR surveillance, the context-specific professional and technical inputs from relevant stakeholders, as well as their knowledge evaluation, acceptance level, and perceived commitment in future program implementation, are vital for the successful policy adoption. Therefore, an exploratory qualitative study for understanding the feasibility, acceptability, and perceptions of the DHIS2 among local OH-AMR stakeholders in Jimma, Ethiopia, has been conducted, which will be translated into many countries. To our knowledge, this is the first study of its kind in the global context for OH-AMR. Taking the implementation barriers into consideration and solving them, DHIS2 can be the most user-friendly and acceptable interoperable platform in resource-poor settings. Piloting at the local level and gradual scaling at the national level with enhanced alignment among OH disciplines followed by adding in the national AMR action plan and health policy should be prioritised for the AMR containment. We believe that this study findings will act as the guiding principles for integrated OH-AMR surveillance policy draughting and the monitoring and evaluation perspective of the whole program.

Sónia Teixeira, Renata Pacheco
INESC TEC–Institutefor Systems and Computer Engineering, Science and Technology, Portugal
Centre for Applied Ecology Baeta Neves, Portugal

PARTICIPATORY METHODS IN CHINA DISCIPLINE ASSESSMENT

ABSTRACT

As human societies developed and their relations became more complex, the need for structures, rules, and compromises arose. Public policies are one of the cornerstones of modern democracies and provide the basis for peaceful and fair coexistence among the citizens of a given society. Evaluation is a crucial process that provides valuable insights into the effectiveness and efficiency of various programs, projects, or interventions. Simultaneously, the evaluation of policies holds the capacity to enhance cost-effectiveness, accountability, and transparency. It also bestows legitimacy upon the utilisation of public funds and resources by furnishing citizens and other stakeholders with insights into whether public decisions, encompassing budget allocation and enacting new laws and regulations, yield the anticipated outcomes (OECD, 2018). The true impact of evaluation lies not only in the assessment itself but in the subsequent utilisation of its results, which becomes key to ensuring sustainability, social balance, and justice in democratic societies.

Several factors contribute to successfully integrating evaluation results into decision-making processes, policy development, and overall improvement initiatives. Artificial intelligence (AI) has been playing an increasingly important role in evaluation, offering a range of tools and techniques for analysing the impact of policy decisions in various areas, from economics to public health. AI's ability to process large volumes of data, identify patterns, and predict outcomes makes it a valuable tool for policymakers.

However, there are significant challenges to be faced in the use of AI in policy evaluation. One of the main challenges is ensuring the transparency and interpretability of AI models, especially in issues that directly affect people's lives. Additionally, ethical issues, such as fairness and data privacy, cannot be overlooked. Another challenge is the constant need to update AI. models to keep pace with changes in society and politics. Policies and their consequences constantly evolve, requiring agile adaptation of AI models to ensure they provide accurate and relevant insights.

The present study aims to answer the following research questions:

RQ1: In what fields are AI methods or tools used for policy evaluation?

RQ2: What type of impact (e.g., cost reduction) does the utilisation of these methods have?

RQ3: Which countries report the highest use of AI methods or tools for policy evaluation?

RQ4: If they exist, what are the cases where a responsible AI perspective is addressed in policy evaluation?

The present study contributes to the scientific understanding of the utilisation of artificial intelligence (AI) tools in policy evaluation. Addressing research questions focused on how AI tools have been employed for policy evaluation and identifying the type of impact will contribute to characterising the landscape of AI applications in policy evaluation. The present work aims to characterise AI tools employed for policy evaluation. This characterisation will be done according to the sector/activity the policy is focused on, its objectives, AI tools/method employed, and results of the experience.

A literature review was conducted to achieve this paper's objective. To gather the best scientific knowledge and experience available on the use of AI for policy evaluation, only peer-reviewed papers from Scopus and Web of Science databases will be analyzed. This review will focus on the AI methods and tools employed, not the paper's overall quality or the suitability of the chosen approaches.

The initial literature search rendered results showing AI tools being employed in various contexts, from construction to health, including e-governance. The preliminary results suggest that AI tools are being explored in many scenarios. This review effort will help guide future research initiatives on evaluating public policies through the use of AI tools. Practitioners can use the results of the present work in the process of choosing the best AI tool to evaluate the policies they are studying. Furthermore, policymakers can also benefit from the results of the present work in understanding the potential for AI in evaluating and informing future policy decisions and reforms.

The main limitations of the present work are linked to its restriction to policy evaluations published in peer-reviewed journals. Other potentially methodologically sound policyevaluation approaches that were not published within the academic context might exist.Furthermore, the topic of AI is one of complexity, and it is evolving extremely fast, which may lead to very different outcomes in the future. To the best of our knowledge, this will be the first work that explores such a topic.

Dr. Filip Zielinski
Centrum für Soziale Investitionen und Innovationen, Universitat Heidelberg, Germany

EVALUATING IMPACT NETWORKS WITH QUIP

ABSTRACT

Impact Networks are a specific type of network that brings together people and organizations based on a common purpose to address social and environmental challenges (Ehrlichman 2021: x). “As a powerful and flexible organizing system that can span regions, organizations, and silos of all kinds, impact networks underlie some of the most impressive and large-scale efforts to create change across the globe” (Ehrlichman 2021:x.). In addition to this assumed capacity to support socio-ecological transformations, their potential to “span silos” makes them particularly relevant from the perspective of innovation and innovation research (Metzger et al. 2012), since they offer a means to enable novel solutions, including services, products and other tangible and intangible objects of innovation, as well as Social Innovation specifically (Howaldt 2019).

At the same time, the impacts of networks in general, and of impact networks specifically, are notoriously difficult to evaluate, since evaluation methodologies are mostly tailored to organizations with their relatively hierarchical and centralized structure. The advantage of networks, i.e. their decentralized structure that enables connections and exchanges in-between nodes, without necessarily involving one and the same central node, poses a challenge, at the same time, to impact evaluation, since it is more difficult to generate relevant processual data, as compared to centralized organizations. The result is that there are relatively little impact evaluations available publicly that concern impacts of impact networks (e.g. Saari 2011, even if they are widely used as a means to deal with complex issues forming a substantial ecosystem of numerous bottom-up, grass-root initiatives that have develop into impact networks, and top-down networks that have been set up by public or private institutions and funders.

Arguably, the representatives and members of impact networks, their external beneficiaries and all potentially affected groups as well as those who fund and support such networks, should be interested in assessing the actual positive and negative, intended and unintended, short-term and long-term, direct and indirect impacts on people, society and the planet. All the more, since such social and/or environmental impacts are naturally part of the Theory of Change of this type of organization.

Evaluation research and practice has brought about a variety of methods and tools as well as conceptual frameworks to enable organizations or external evaluators to assess impacts. For practical reasons (decentral structure, as outlined above) and also to make use of the advantages or participative

evaluation, an impact evaluation of impact networks needs to involve the nodes (people, organizations) of the network when gathering relevant data. After a review of available quant-led and qual-led approaches (Copestake 2024), the method of the Qualitative Impact Protocol developed by the University of Bath, UK, and put into practice by BSDR (2020) in various evaluation projects over the past years – some of them, interestingly, in combination with RCTs – seems like a particularly promising option for this type of application. As far as the author is aware, and according to BSDR, the QuIP method has been used to evaluate a network of organizations only once so far. In an exploratory study, funded by AEC, a European network of Higher Music Education Institutions, the author has tested the approach making use of an annual congress of the network to conduct interviews with representatives of the network's members, which were then analysed using the QuIP methodology. The results of the evaluation have been reported back to AEC in order to inform and support their strategy development, particularly to check findings against their current Theory of Change. The presentation and paper will discuss the advantages and disadvantages of the QuIP methodology and illustrate them using the pilot study outlined above, including the tools that BDSR has developed to visualise the causal pathways including quantitative date and to connect these causal links with the relevant interview contents. Current developments include using AI to conduct QuIP interviews via chatbots. This additional aspect should be of interest since conference strand 5 deals with digital tools, AI and big data. The advantages of QuIP for impact evaluations of Impact Networks include that it provides a necessary level of standardization of a largely qualitative approach while allowing to account for high complexity of context; it allows to gain valuable insights into the issue of attribution of outcomes without relying on control groups; it gives a voice to beneficiaries and takes relatively less time and resources than experimental approaches. On the downside, it still requires substantial know-how and resources, compared to e.g. a network member survey; it performs well on outcome domains that involve members directly, but probably less in those that do not, specifically advocacy; it is mostly founded in the theory-based conceptual framework of impact evaluation and while it allows for some quantification, it does not offer counterfactual comparisons in the strict sense, as understood by the variance-based conceptual framework of impact evaluation.

REFERENCES:

- BSDR (2020). QuIP briefing paper. https://bathsdr.org/wp-content/uploads/2020/10/QUIP-briefing-paperOctober-2020.pdf

- Copestake, J. (2024). Mixed-methods impact evaluation in international development practice: distinguishing between quant-led and qual-led models. Journal of Development Effectiveness, 1-20.
- Ehrlichman, D. (2021). Impact networks: Create connection, spark collaboration, and catalyze systemic change. Berrett-Koehler Publishers.
- Howaldt, J. (2019). Soziale Innovation im Fokus nachhaltiger Entwicklung–Die Bedeutung von Kooperationen und Netzwerken für den Erfolg sozialer Innovationen. Netzwerke und soziale Innovationen: Lˆsungsans ‰ tze f¸r gesellschaftliche Herausforderungen?, 13-30.
- Metzger, F. M., Berwing, S., Armbr¸ster, T., & Oberg, A. (2012). Koordinationsmechanismen und Innovativitaet von Netzwerken: eine empirische Analyse. Schmalenbachs Zeitschrift fuer betriebswirtschaftliche Forschung, 64, 428-455.
- Saari, E., & Kallio, K. (2011). Developmental Impact Evaluation for Facilitating Learning in Innovation Networks. American Journal of Evaluation, 32(2), 227-245.

Peter Kolarz
Technopolis, United Kingdom
INSIDE THE FUNDING PROCESS

ABSTRACT
Research funding schemes may have several different aims. These include (but are not limited to), funding basic exploratory research, research that is societally relevant (including to specific societal challenges), research that has particular promise for industrial application, research that is highly innovative and breaks with established paradigms, research that presents new interdisciplinary or transdisciplinary perspectives, or research that seeks to foster collaboration between previously unconnected individuals.
Especially in cases of schemes with multiple aims – and consequently, multiple assessment criteria – process evaluations of funding schemes face the challenge of having to assess to what extent application selection processes consider and reward these various aspects of proposed research projects. Put differently: how well do the review processes actually operationalise the aims of the funding instrument?
Moreover, recent years have seen a proliferation of modifications to the long-established standard assessment process for research grants (typically external peer review followed by expert panel review using standard criteria around research quality and feasibility). These include short preproposals, inclusion of non-academic reviewers or panellists, in-person presentations. Often, funders have introduced such modifications in order to ensure assessment of a range of different aspects of applications submitted to a scheme.
For evaluators, it therefore becomes important to assess, how different assessment criteria feature in different parts of application assessment processes and ultimately, whether a funding process therefore actually rewards and prioritises the assessment criteria that reflect the aims of the funding instrument.
With this objective, we used generative AI to analyse peer reviewers' reports on applications submitted as part of a process evaluation of the Austrian FWF's Emerging Fields (EF) programme. The EF programme aims to fund collaborative research teams to conduct pioneering basic research that departs from established approaches. It aims to give researchers the opportunity to pursue particularly innovative, original, or high-risk ideas.
The EF scheme has two separate written application assessment stages: a short outline-proposal stage, followed by a full application review stage. Background research on the scheme's design led us to a hypothesis that reviews in the first of these two stages should emphasise and reward innovative potential and novelty of the proposed project ideas, while reviews in the second stage should place a greater emphasis on scientific quality of the research plans.
Prominent international scientists reviewed the stage-1 synopses and stage-2 full proposals for the EF programme and produced evaluation documents detailing their judgements. In total, we had access to 140 records: 87 peer-reviews of synopses and 53 reviews of full proposals. Given the number, heterogeneity and complexity of the review documents,

generative AI was particularly useful in facilitating a systematic assessment. The multidisciplinary nature of the EF programme and the fact that EF applications are at the frontier of science make the review documents far from digestible for a general audience. They are heavy on scientific jargon and technical details, which are hard to understand for someone without deep expertise in each topic.

We used a rich and comprehensive generative AI model that managed to navigate these technical details to find relevant individual insights about each review document and stylised facts about the selection process in its two stages.

The model was OPENAI's latest GPT4 large language model accessed programmatically via a dedicated API. This access mode ensures the privacy and confidentiality of the underlying data. Also, this mode of access enabled us to explore the capabilities of the GPT4 model in large-scale automation (querying all the documents programmatically without manually inputting and querying each review document individually).

To test our hypothesis, we performed topic detection, sentiment and priority detection analyses, exploring the text of the review documents from both the first and second assessment stages with generative AI. The topic detection consisted of distinguishing the parts of the reviews focusing on novelty, risk, scientific quality and team suitability dimensions. The sentiment analysis assessed whether the reviewers were positive, negative or neutral on their considerations about each topic. Finally, the priority detection measured the text length or number of words dedicated to each topic as a proxy for how much effort/time the reviewers devoted to each dimension.

Moreover, to ensure accountability and transparency of the model's outputs we requested the generation explanations for each resulting score accompanied with quotes from the original reports for human cross-validation. Next, we transformed the topic-level sentiment and priority indicators into quantitative scores. For the sentiment variables, we generate one sentiment variable per topic/dimension and for each document.

Our results provide evidence in favour of our main hypotheses, most notably with novelty having a significantly higher priority at stage-1 and scientific quality having a significantly higher priority at stage-2.

The analysis of priorities confirms that at the synopses review stage, reviewers focus more on the novelty dimension and less on considerations about scientific quality. Our ranking indicator ranges from 0 to 3, with novelty scoring, on average, 2.26 for stage-1 synopses and only 1.25 for stage-2 full proposals. The 'risk' and 'team' dimensions also receive more attention at the synopses stage than proposals, but with less pronounced differences. The differences in terms of dedication to reviewing scientific quality are more noticeable. This dimension has the lowest rank at stage-1, scoring on average 0.61. In contrast, scientific quality considerations rank the highest at stage-2, scoring 2.19. When assessing the breakdown of priority ranks per successful and unsuccessful applications, the pattern remains unchanged regarding differences in focus. However, the breakdown reveals that in the case of successful proposals, reviewers dedicate, on average, slightly more effort to novelty considerations in comparison to unsuccessful applications. In contrast, for unsuccessful applications, reviewers allocate more effort to review scientific considerations, potentially with a more comprehensive review of existing pitfalls.

We did not conduct this analysis in isolation. Our process evaluation of the EF programme also included surveys of various stakeholder groups, expert interviews, and observation of review meetings. We also produced novelty indicators suggesting that, overall, applications to the EF programme reveal high levels of novelty when compared with other FWF programmes. Insights form these other method components helped to formulate our hypothesis for the analysis of reviews and also provided various forms of triangulation, providing additional confidence in our findings.

This work demonstrates the possibilities and benefits of using generative AI in the analysis of peer review reports. The detailed texts of peer reviewers are generally an under-used resource in funding scheme process evaluations (in part due to their technical and thematic complexity) and we are able to suggest a viable way of making more use of such reviews to help assess whether assessment processes are working as they should.

In light of the many different aims that various funding instruments may pursue, and the increasing range of different assessment processes used to select funded applications, our approach provides an innovative way of gaining insight into the operationalisation of assessment criteria in a range of different contexts. This may go some way towards better understanding what kinds of funding processes work best in various situations.

Ioanna Grypari, Dimitris Pappas, Haris Papageorgiou
Opix & Athena Research Center, Greek

IADVANCING RESEARCH IMPACT EVALUATION IN THE DIGITAL ERA: INSIGHTS FROM EUFUNDED RARE DISEASE PROJECTS

ABSTRACT

Acknowledgement: This work was partially funded by the European Union through the Horizon 2020 projects Data4Impact (Grant Agreement ID: 77053) and IntelComp (Grant Agreement ID: 101004870). The views and opinions expressed herein are solely those of the authors and do not necessarily reflect those of the European Union or the relevant granting authorities. Neither the European Union nor the granting authorities can be held responsible for them. Special thanks to Christina Kyriakopoulou for her meticulous curation of the list of rare disease projects and other project outputs, and for offering her valuable insights in the earlier stages of this work.

One of the principal challenges in public research funding is its capacity to effectively evaluate when and if the generated outcomes can impact society. This challenge primarily stems from the difficulty in tracking the progression of research developments that typically span several years or even decades before yielding mature societal solutions or economic products. Recognizing that valuable insights cannot be derived from a single source alone, this study introduces a comprehensive methodology that integrates a network of cross-linked data sources, enhancing policy-making insights. This approach is exemplified through a case study in the domain of rare diseases.

To manage the vast amount of data generated from diverse research activities and the increasing size of databases, our methodology employs sophisticated analytical tools and state-of-the-art document analysis techniques. These include advanced artificial intelligence tools complemented by human/expert-in-the-loop steps where necessary. This integration is crucial for understanding the evolution of research development and ensuring that policy interventions are well-targeted and timely.

Moreover, the technological capabilities we currently possess to store, process, link, and analyse big data are leveraged to serve policymakers. This enables the monitoring and evaluation of policy interventions, helps set their agenda, and understand the impact pathways of different types of research and innovation activities. Additionally, we propose the development of innovative indicators to better track the evolution of research developments, achieved by extracting and synthesizing information from both structured and unstructured data sources. This transformation incorporates a qualitative dimension into our analysis, enriching traditional metrics with a deeper contextual understanding, thereby enhancing the capacity of public research funding to assess and respond to the long-term outcomes of its investments effectively. This approach integrates a 'context dimension' that transcends the limitations and pitfalls associated with simple terminology (e.g., keywords) and traditional methods of associating and linking data sources, thus providing a more nuanced and comprehensive evaluation framework.

Our analysis draws from a wide range of data sources associated with the FP7 and H2020 EU-funded programs, with a specific focus on rare diseases. These sources include project data, publications, patents, company websites, clinical guidelines, clinical trials, and drug development efforts. By integrating these diverse data types, we enable a robust analysis of the research lifecycle and its broader societal and economic implications.

The foundation of our approach involves the strategic application of big data technologies to manage and analyse our extensive and varied datasets. We utilize advanced data analytics tools, including state-of-the-art Natural Language Processing techniques such as knowledge graphs and ensemble algorithms. These tools are essential for extracting, processing, and synthesizing information from different data sources, producing accurate, timely, and detailed insights from research inception to impact. Within this framework, we employ novel machine-learning methods to classify scientific articles according to Field of Science (FOS). This allows us to analyse the distribution and focus of research activities within EU-funded projects, aiding in identifying research trends, gaps, and interdisciplinary connections. This classification supports strategic planning and funding allocation for future research initiatives.

Further, we implement an ensemble classifier that combines deep learning, topic modelling, and handcrafted rules to categorize scientific literature according to the United Nations' Sustainable Development Goals (SDGs). This classification helps evaluate the alignment of EU-funded research with broader societal objectives, ensuring that efforts contribute to global priorities like health and sustainable development.

Our methodology also includes the creation of a citation graph that incorporates patents, clinical trials and guidelines, establishing links between EU-funded research outputs and other health-related outcomes. By analysing this database with techniques like graph analysis and topic detection, we can trace how EU-funded scientific literature contributes over time within specific scientific domains. Patents and clinical trials included in the graph provide insights into the commercialization potential of research and its translation into innovative products and technologies. This linkage demonstrates the progression from basic research to clinical applications and patient care.

We also assess changes in industry engagement by calculating companies' uptake scores to determine if they continue working on project topics years later or shift focus. This insight helps understand long-term impacts of research funding and guides strategic resource allocation in response to evolving market needs.

Lastly, we enhance our evaluation framework by linking mentions of diseases in scientific literature to the International Classification of Diseases (ICD) ontology. This linkage not only aligns EU-funded research with international health standards but also allows for systematic evaluation of research relevance and impact against global health challenges, enabling more targeted and effective policy interventions.

Human Validation and AI Integration: A key feature of our methodology is the integration of human-in-the-loop processes, where domain experts provide essential validation and refinement of the data and its analysis. This human validation is critical to ensure the accuracy and relevance of the insights generated. Alongside human expertise, we leverage high-performance computing and advanced AI techniques to handle dynamic, multilingual, and heterogeneous data, enhancing the system's capacity for agile decision-making.

Transparency and Replicability: Our system is designed to be open, transparent, and replicable, fostering trust and sustainability in research evaluation. We prioritize the use of Open and FAIR data4 alongside methodologies that are openly accessible and transparent, ensuring that our findings can be replicated and verified by other researchers and policymakers.

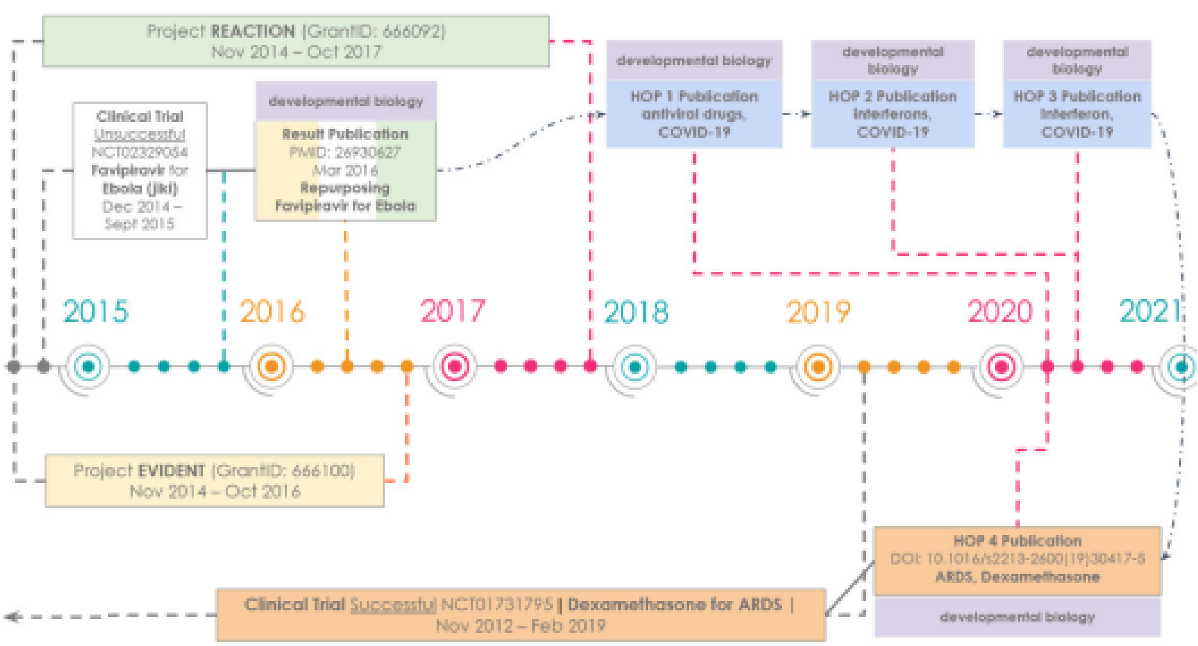
Rare diseases are defined as conditions that affect no more than 1 person in 2,000 within the European Union. Currently, between 6,000 and 8,000 distinct rare diseases are recognized, affecting an estimated 36 million people across the EU. The complexity and diversity of these conditions necessitate a collaborative approach at both European and international levels to advance treatment and diagnosis.

Rare diseases pose specific challenges due to their low prevalence and the significant variability in their manifestations. Nearly 70% of these diseases begin in childhood, which complicates diagnosis, treatment, and care processes. In response, the EU has supported extensive research through programs like FP7 and Horizon 2020, investing over €2.9 billion into more than 600 projects focused on rare diseases from 2014 to 2020. This investment is designed to foster collaboration, enhance diagnostics, and improve treatment effectiveness. Our study analyses data from two EU-funded research programs, FP7 and Horizon 2020. This selection reflects the need to examine the medium to long-term development times of research outputs and their societal impact, juxtaposing policy time against R&D time. Additionally, our analysis considers the impacts of the COVID-19 pandemic, which necessitated the repurposing of some research directions and influenced research outputs and potential impacts in subsequent years. A critical aspect of our approach is the establishment of a robust portfolio of projects that accurately represents the EU's investment in rare disease research. This process involves both automated AI techniques and manual validation by human experts to ensure the relevance and accuracy of the data.

Our analysis sheds light on critical insights derived from the examination of rare disease research, which have significantly impacted policy-making and research direction.

- Longitudinal Perspective: By looking back over an extended period, we capture the maturation of societal impacts and the environment that drove these outputs. This historical lens provides valuable context for understanding the evolution of research outcomes.
- Context-Driven Analysis: Contextual analysis, reliant on both structured and unstructured data sources, is crucial for informing policy-making effectively. This approach ensures that evaluations are sensitive to the nuanced complexities of the research landscape.
- Deeper Analytical Insights: While aggregate statistical analysis offers valuable insights, deeper analysis of combined sources is essential for piloting future policy-driven decisions. This nuanced understanding is critical for strategic orientation and evidence-based policymaking.
- Building Trust and Skills: Our study emphasizes the importance of building trust and skills in using data-driven insights to support policymaking. Automation of certain tasks enhances scalability and flexibility, enabling the development of new evidence for informed decision-making.
- Complementary Intervention Approaches: Our work contributes to the development of interventions based on advanced data analytics methods, complementing traditional approaches. This is particularly relevant given the diverse and expansive nature of the data landscape.

A Clinical Trial Use Case:



This timeline depicts the interconnected journey from initial research projects to a successful clinical trial outcome. Each box represents a research project, clinical trial or publication, with arrows indicating citations or references between them. Through a series of four hops, represented by citations, the pathway leads from an unsuccessful project-funded clinical trial to a successful clinical trial. The topics in each box highlight the areas of research contributing to the eventual outcome. This visualization demonstrates the intricate connections that can be automatically mapped on a large scale, revealing pathways that might otherwise be overlooked. Moreover, it highlights the need for cautious interpretation, as the rarity of this pathway—one in ten thousand—emphasizes the importance of rigorous evaluation methodologies.

Our study on rare diseases within FP7 and Horizon 2020 showcases a possible effective integration of digital evaluation tools and AI in research assessment. Through the utilization of diverse data sources, such as scientific publications, company websites and clinical trials, we offer nuanced insights crucial for evidence-based policy-making.
Key findings underscore the significance of contextual, data-driven approaches and the potential for automation and scalability in research evaluation. Moreover, the challenges presented by the COVID19 pandemic highlight the adaptability of science in responding to evolving priorities and demands.
Furthermore, our study sheds light on the interplay between policy-makers' missions and interventions, emphasizing the importance of aligning research efforts with overarching societal goals. By augmenting human decision-making with AI-driven insights (policy intelligence), policymakers can observe patterns, identify emerging trends, and extract lessons learned to inform more effective and efficient policy-making processes. This approach facilitates a forward-looking perspective, enabling policymakers to anticipate future challenges and proactively shape policies that address evolving societal needs.

Gabriel Okasa, Michaela Strinzel, Anne Jorstad, Katrin Milzow, Matthias Egger, Alberto de León, Stefan Müller
Swiss National Science Foundation, Switzerland
Swiss National Science Foundation, Switzerland
Swiss National Science Foundation, Switzerland
Swiss National Science Foundation, Switzerland
Swiss National Science Foundation, Switzerland
University College Dublin, Ireland
University College Dublin, Ireland

UNDERSTANDING GRANT PEER REVIEW USING MACHINE LEARNING

ABSTRACT

Peer review lies at the heart of evaluating scientific manuscripts, grant project proposals, and ultimately scientific progress (Steiner Davis et al. 2020). Despite its central importance for academic research and innovation, systematic analyses of peer review reports are challenging. Peer review reports are often confidential, and the publication of peer review reports alongside published research is a rather recent and limited phenomenon. Gaining systematic insights into grant peer review reports can contribute to a better understanding on what basis funding decisions are made, decisions that can mean the make or break of groundbreaking discoveries and the career of researchers. This study combines qualitative coding and state-of-the-art transformer-based machine-learning approaches to assess characteristics of more than 36'000 grant peer review reports submitted to the Swiss National Science Foundation in the largest funding scheme Project Funding between 2017 and 2023. The approach suggested in this study can help funders to understand how peer reviewers judge proposals and if they adhere to their rules and guidelines. This can inform modifications of the review process and can provide feedback for improving reviewer training.

We first developed an annotation codebook capturing characteristics of grant peer review reports on sentence level. Most of these categories directly mirror the Swiss National Science Foundation's (SNSF) guidelines. We conducted eleven separate manual annotation exercises. During these exercises, we developed and improved a detailed annotation codebook and instructions. The final codebook consists of 14 distinct categories, each aligned with specific criteria and principles outlined in the SNSF requirements. Broadly speaking, our categories capture one or more of the following aspects:

1. Does the review refer to the researcher(s) or to the project?
2. Which SNSF review criteria does the reviewer apply?
3. Is there a mention of an impact beyond academia?
4. Does the review make a positive statement, and/or a negative statement?
5. Is a rationale for the positive and negative statements provided?

The final codebook served as a crucial reference guide for the four annotators with a scientific background involved in the study. They meticulously annotated each sentence of

a sample of 3,000 sentences in all 14 categories, adhering to the definitions provided in the codebook. The first set of 2,000 sentences were sampled randomly from the entire corpus of English reviews. A stratified sampling strategy was used for the remaining 1,000 sentences. More precisely, we sampled 200 sentences from each of the five questions ofSNSF reviews to obtain more relevant sentences for under-represented categories, A robust validation strategy was implemented to address the potential subjectivity inherent in individual annotation decisions. Identifying and annotating peer review characteristics is a challenging task, and one evaluation of each sentence is not sufficient. For this reason, we randomly allocate each sentence to three of the four coders. We aggregate the coding decisions by the three coders using majority agreement. If at least two of the three coders identified that a certain characteristic was present in the sentence, we treat it as an example of the category. If none or only one coder identified a certain characteristic, we do not treat the sentence as an example of the category. This consensus-driven approach seeks to mitigate the impact of individual coder bias (e.g., Benoit et al. 2016), thereby enhancing the reliability of the annotated dataset.

As part of validation measures, we first regard levels of agreement between the coders, which ranges from 62 percent to 98 percent, with an average of 80 percent agreement across all categories. Second, we examine the percentage of sentences per category annotated with 'yes' or 'no' by one, two, or three coders. Many categories tend to be imbalanced, meaning that only few sentences have been labelled as 'yes' by all three coders. In addition, a comprehensive analysis of agreements and disagreements between coders was conducted after each annotation round. This process facilitated a nuanced understanding of annotation discrepancies, enabling continuous refinement of the coding guidelines, feedback to the annotators, and improving overall coding consistency.

This systematic and collaborative annotation process establishes a robust foundation for subsequent analyses, including fine-tuning the transformer model for the classification task. The attention to detail in the annotation phase contributes to the validity of the classification, laying the groundwork for all subsequent analyses.

We seek to obtain predictions for the presence or absence of the pre-defined categories for all sentences contained in our sample of the SNSF grant peer reviews. Given the overwhelming amount of text data available, it is not feasible, and possibly not even reliable, to obtain the labels via human annotation. Therefore, we use the human annotated sample of 3,000 sentences to train machine learning models that predict the labels for the remaining sentences in our dataset. For this purpose, we rely on a pre-trained language model and fine-tune it based on our set of annotated sentences.

We follow the recent advances in deep learning for natural language processing and make use of a pre-trained transformer model (Vaswani et al. 2017, Tunstall et al. 2022). We use

the so-called BERT model (Devlin et al. 2018), i.e. deep bidirectional transformer. Given the specific scientific domain in our sample of grant peer reviews we use a version of the BERT model that used additional scientific texts for pre-training and which was further augmented by citation graph, i.e. the SPECTER model (Cohan et al. 2020). More precisely, we use SPECTER2 (Singh et al. 2022), an updated version of the original SPECTER model.
The pre-trained SPECTER2 model as such can be used off-the-shelf for extracting the text representation (Cohan et al. 2020). To fully leverage the potential of the pre-trained model, we not only extract the pre-trained embeddings, but directly fine-tune SPECTER2 for a binary classification task using the annotated set of 3,000 sentences. As such, we add an additional linear output classification layer on top of the pre-trained model, which we fully train, whereas the parameters of the pre-trained model layers are fine-tuned. We divide the labelled data into training (90%) and test sets (10%). We use the training set of 2,700 sentences to fine-tune the model and the test set of 300 sentences to evaluate the prediction accuracy.
Adjusting the hyperparameters while fine-tuning a model allows for optimal prediction performance, i.e. to not over-fit on the training data and thus compromise the generalization of the model (Hastie et al. 2009). For this reason, we perform a 5-fold stratified crossvalidation schema for the hyperparameter tuning. Given the best-performing hyperparameter setting, we fine-tune the SPECTER2 model on the full set of 2,700 sentences from the training set and evaluate the prediction accuracy on the not yet touched test set of 300 sentences to get an unbiased evaluation of the fine-tuned model. We perform this fine-tuning procedure for all 14 categories. We evaluate the prediction performance based on the standard classification accuracy, precision, recall as well as the F1-score (macro-average of both class labels) to reflect the imbalance in the class labels.

The preliminary results reveal a large heterogeneity in the classification performance across categories. Additionally, there is a large discrepancy between the classical accuracy measure and our preferred measure, the F1-score, which shows that the class imbalances play a substantial role. Overall, the F1-score ranges from almost 58% to over 94%, with an average of 82% across all 14 categories. Interestingly, lower F1-scores are associated with both a lower share of sentences with the label itself as well as with a lower coder agreement. This sheds light on the difficulties of accurate classification for categories that appear infrequently and are intrinsically difficult to code for human annotators.
While the empirical analysis is still at an early stage, we have assessed the validity of the classification in various ways. Moving beyond the performance metrics, we assess the face validity in two ways. First, building on recent advances in explainable machine learning

(Molnar, 2020), we apply the method of integrated gradients (Sundararajan, Taly, Yan, 2017; Janizek, Sturmfels, Lee, 2021) to uncover which words contribute the most for the class predictions. Second, following Severin et al. (2023) we also conduct a keyness analysis to identify terms that are predictive of each category.

Finally, we apply the fine-tuned models to predict the categories for all sentences within the corpus of the grant peer review reports and aggregate the prediction results onto a review level. Initial results reveal that some categories appear very frequently across reviews, that grant peer reviews tend to be more positive than negative, and that some of the categories are difficult to classify for human coders and cannot be accurately predicted by our transformer-based models. We also find substantial differences in grant peer review characteristics across disciplines and reviewers' backgrounds, including gender.

Our study has important implications for researchers and funders. From a technical perspective, the initial results show that natural language processing and machine learning methods can be successfully applied to grant peer review reports. Our findings will also contribute to a better understanding of characteristics of grant peer review reports and may help funding agencies to assess existing practices, which could enhance the review process. Finally, researchers will gain a better understanding of the grant peer review process, potentially increasing their engagement and compliance with review guidelines.

REFERENCES:

- Benoit K, Conway D, Lauderdale BE, Laver M, and Mikhaylov S (2016). Crowd-sourced Text Analysis: Reproducible and Agile Production of Political Data. American Political Science Review, 278–295.
- Cohan, A., Feldman, S., Beltagy, I., Downey, D., and Weld, D. S. (2020). Specter: Document-level representation learning using citation-informed transformers. arXiv preprint arXiv:2004.07180.
- Devlin, J., Chang, M., Lee, K., and Toutanova, K. (2018). BERT: pre-training of deep bidirectional transformers for language understanding. CoRR, abs/1810.04805.
- Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction, volume 2. Springer.
- Janizek, J. D., Sturmfels, P., & Lee, S. I. (2021). Explaining explanations: Axiomatic feature interactions for deep networks. Journal of Machine Learning Research, 1–54.
- Molnar, C. (2020). Interpretable machine learning. Lulu.com.
- Severin A, Strinzel M, Egger M, Barros T, Sokolov A, Mouatt JV, et al. (2023) Relationship between Journal Impact Factor and the Thoroughness and Helpfulness of Peer Reviews. PLoS Biology, e3002238.

- Singh, A., D'Arcy, M., Cohan, A., Downey, D., & Feldman, S. (2022). Scirepeval: A multiformat benchmark for scientific document representations. arXiv preprint arXiv:2211.13308.
- Steiner Davis MLE, Conner T.R, Miller-Bains K, and Shapard L (2020) What makes an effective grants peer reviewer? An exploratory study of the necessary skills. PLoS ONE, e0232327.
- Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. In International conference on machine learning (pp. 3319-3328). PMLR.
- Tunstall, L., Von Werra, L., & Wolf, T. (2022). Natural language processing with transformers: Building language applications with Hugging Face. Beijing: O'Reilly.
- Van den Besselaar, P., Sandström, U., & Schiffbaenker, H. (2018). Studying grant decisionmaking: a linguistic analysis of review reports. Scientometrics, 313–329.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems.

Stefanie Schuerz, Charlotte d'Elloy, Michael Dinges
Zentrum für Soziale Innovation GmbH, Austria
Technopolis Forschungs- und Beratungsgesellschaft m.b.H., Austria
Austrian Institute of Technology GmbH, Austria

THE INTERSECTION OF AI AND RTI (POLICY) EVALUATION

ABSTRACT

Our draft paper "The Intersection of AI and RTI (Policy) Evaluation: Principles and Considerations" presents a comprehensive exploration of the integration and implications of artificial intelligence (AI) in the field of research, technology and innovation (RTI) policy evaluation, specifically in the Austrian context. The paper is structured to first provide working definitions of AI, followed by a detailed discussion of key principles and considerations for using AI in evaluation processes, and concludes with an overview of current practices and future prospects.

The paper begins with a discussion of the transformative impact of AI in organisational processes, improving productivity and supporting data analysis. AI, particularly Generative AI, is acknowledged for its potential beyond mere technological hype, suggesting its prospective essential role in evaluation work. The authors highlight the importance of using AI responsibly to maintain the quality and rigour of evaluation practice. They also stress the need for clear disclosure and standards when integrating AI into professional work to ensure informed decision-making and integration.

The methodology section details the systematic approach taken by the fteval community, including literature reviews and surveys, to gather insights into the adoption and impact of AI in the Austrian RTI evaluation community. This foundational work helps to frame the subsequent discussion of principles and best practices for the use of AI.

The document elaborates on several key principles vital for the ethical and effective use of AI:

- Respect for Human Autonomy and Responsibility: Even when using AI systems, human users remain responsible for the accuracy, transparency, and accountability of outputs created with AI support and shared in a professional context. While AI applications can be a valuable tool for R&I, they do not replace critical thinking, human expertise, and rigorous scientific methodology. Any outputs need to be checked for overall correctness, plausibility, as well as any ethical aspects that might be relevant, such as biases inscribed in the text. As such, human oversight and responsibility are essential to AI integration.
- Transparency and Explainability: The application of AI systems for R&I evaluation, and the purposes for this application, need to be disclosed, especially 1) when AI systems are being used for analytical purposes; 2) when reproducibility of results is required; and 3) when the output created by AI is likely to have an impact on decision-making processes. Explainability is about enabling people that are affected by the outcome of an AI system to understand how this outcome was arrived at. This entails the provision

- of easy-to-understand information that enables anyone adversely affected by the AI system to challenge the outcome. This clarity about AI use in evaluative processes ensures accountability. -
- Prevention of Harm: AI systems should neither cause nor exacerbate harm, be it collective or individual harm, or otherwise adversely affect human beings. As a part of this, human dignity and mental and physical integrity of individuals and communities need to be protected. Potential risks are exacerbated whenever AI systems are used for decision-making processes. Thus, human verification is a must for process-sensitive activities such as writing manuscripts and data scripts, peer review, proposal evaluation, and so on.
- Fairness and Non-Discrimination: Fairness has a substantive and a procedural dimension. The substantive dimension covers a general commitment to fairness such as equally and justly sharing both benefits and costs of AI, promoting social justice and minimising discriminatory applications and outcomes of AI. The procedural dimension is more operational and entails the ability to contest and seek effective redress against decisions made by AI systems and by the humans operating them.
- Privacy Assurance: Privacy is a fundamental right particularly affected by AI systems. It necessitates adequate data governance that covers the quality and integrity of the data used, its relevance within the domain in which the AI systems will be deployed, its access protocols, as well as the capability to process data in a manner that protects privacy throughout the entire AI lifecycle. The General Data Protection Regulation is also relevant when employing AI systems, and lays out the need to provide easily accessible and understandable information about our data collection practices and the processing practices we employ, including our reasons for data collection and processing. As such, the GDPR lays out existing data privacy standards, particularly in handling personal and sensitive data, which need to be followed.

The discussion moves on to the current use of AI in RTI evaluation, noting a gradual increase in the use of AI for various evaluation tasks, such as data analysis, process automation, and overcoming language barriers. Despite this progress, there are ethical challenges and the potential for AI to impact decision-making and integrity in research.

The paper also explores the potential synergies between proprietary AI systems and different phases of the evaluation process. It outlines specific use cases for AI in RTI (policy) evaluation and identifies how AI can improve the efficiency and effectiveness of evaluations. The paper explores synergies at the intersection of proprietary AI systems and evaluation phases, and discusses how AI can be integrated into different phases of the evaluation process. It points out that AI technologies can support evaluators by providing tools for data collection, analysis and reporting. For example, AI can be used to automate the extraction of relevant data from large datasets, which is particularly valuable in areas where the volume of data can be overwhelming for human analysts. The paper highlights how AI systems can be used to improve the transparency and objectivity of ratings. AI tools can help identify biases and inconsistencies in the evaluation process, improving the fairness and accuracy of project evaluations. In addition, AI can speed up the evaluation process by automating routine tasks, allowing human evaluators to focus on more complex decision-making aspects.

Finally, the paper proposes a self-assessment list for organisations considering the use of proprietary AI systems for evaluation purposes. The list is intended to help evaluators and programme managers assess the suitability of AI tools for their specific needs and contexts. It includes considerations such as the AI system's capabilities, alignment with organisational goals, ethical implications, and compliance with relevant laws and regulations.

Looking to the future, the "AI Guidelines for Evaluation" advocates for continued education and training on AI technologies to ensure their responsible use in evaluation contexts. It highlights the importance of developing robust guidelines and ethical standards that keep pace with technological advances to ensure the integrity and effectiveness of evaluation practices.

The paper concludes by affirming the potential of AI to significantly enhance RTI evaluation processes, but calls for a balanced approach that considers the ethical, legal and social implications of technology adoption. It emphasises the need for ongoing dialogue, policy development and training to fully exploit AI in a manner that is both innovative and responsible.

Carlos Vergara
Technopolis Group, United Kingdom
REF 2028 ENVIRONMENTAL INDICATORS COMMISSION

ABSTRACT
The Future Research Assessment Programme (FRAP) is an initiative by UK and devolved ministers and higher education funding bodies to explore new approaches for assessing research performance in UK higher education. Its initial focus is shaping the next Research Excellence Framework (REF) exercise, set for completion in 2028. REF 2028 aims to enhance the evaluation of research conditions, requiring institutions to submit structured statements supported by outcomes-focused indicators to demonstrate their research strategy effectiveness.

The Research Excellence Framework (REF) is the UK's mechanism for assessing research quality in higher education institutions. It replaced the Research Assessment Exercise in 2014 and was last conducted in 2021. The REF influences the allocation of research funding, ensures accountability for public investment, and provides insights into the health of research in the UK. The assessment process is managed by expert panels composed of senior academics, international members, and research users, evaluating various research outputs and impact case studies.

Research culture and environment are increasingly recognised as critical components of excellent research. The REF 2021 emphasised the need for a consistent, data-driven approach to assess research environments, accounting for 15% of the overall outcome, focusing on vitality and sustainability. There has been growing interest in improving research culture, with various initiatives and reports highlighting the need for better practices to enhance research quality and researcher well-being.

While there is no universally agreed definition of research culture, several frameworks provide a basis for assessment. The Concordats and Agreement review identified domains such as research conduct, staff development, research assessment, and equality, diversity, and inclusion (EDI). Science Europe's values framework includes autonomy, care, collaboration, integrity, and transparency. These elements are often reflected in institutional research strategies and policies.

The overarching aim of this project is to develop and test methods for assessing research culture and environment, identifying indicators for institutional and sector-level evaluations. To meet these objectives, the partnership with Technopolis Group coordinated the delivery of four blocks of work:

- Co-designing methodology with partners and reviewing existing frameworks and studies. This includes an analysis of institutional and unit-level environment statements from REF 2021.

- Consulting the sector to refine the assessment framework and indicators.
- Engaging REF 2028 panels to finalise the methodology.
- Synthesising advice into actionable recommendations.

The project used generative AI to analyse 1,935 environment statements from REF 2021. This analysis involved a search, classification, and question-and-answer approach to identify evidence related to research culture and environment. The use of AI facilitated systematic identification of themes, summarising, and synthesising information to ensure transparency and accountability. The process included developing prompts based on previous research culture frameworks and deploying them using the GPT-4 model. Generative AI was employed to analyse content from the REF 2021 environment statements, identifying themes and evidence related to research culture and environment. This approach provided a detailed, systematic overview of the data available in these statements, enabling the identification of common topics and indicators across institutions.

For example, for the topic of "good research practice and integrity", the AI was prompted to summarise discussions and list indicators and evidence, highlighting transparency, accountability, and governance practices.

Nine prompts were prepared in total for each of the 4 different pillars on research culture and environment. Each of these had follow up prompts to synthesise a list of common topics and indicators. The summaries collected were inspected to derive insights and conclusions regarding the different themes on research culture and environment.

The analysis of REF 2021 submissions revealed several preliminary results related to the key themes:

- Evidence of Research Management: Aligning with proposed indicators, although often lacking specificity.
- Evidence of Research Value: Focused on open access, public engagement, and policy impact, with some overlaps in ethical research practices.
- Support for People: A large volume of evidence related to diverse workforce recruitment, recognition, and career development, although generally less specific than the proposed indicators.
- Engagement with Others: Consistent evidence across panels related to research leadership, with overlaps in workforce diversity and career development areas.

The institution-level analysis indicated alignment with the proposed long list of indicators for REF 2028, suggesting that HEIs are largely familiar with the main themes under consideration. However, there was a noted lack of specificity in the 2021 submissions. Institutions provided similar evidence types as at the unit level, but the precision and specific approaches to measuring phenomena varied.

The REF, in partnership with Technopolis Group, aims to refine and enhance the assessment of research culture and environment for REF 2028 by developing a structured, evidence-based approach. The use of generative AI facilitated a comprehensive analysis of REF 2021 environment statements, identifying key themes and indicators that will inform future assessments. This methodology ensures a consistent and transparent evaluation of research environments, contributing to the overall improvement of research quality and culture in the UK.

Thomas Palfinger
LBG OIS Center, Austria

ARTIFICIAL INTELLIGENCE IS COMING. HOW TO REACT?

ABSTRACT

This contribution is based on the results of a working group of the Austrian Platform for Research and Technology Policy Evaluation (fteval) on Artificial Intelligence (AI). In this context, members of the fteval community discussed and worked on various aspects and topics regarding AI from July 2023 to March 2024. This specific piece focuses on the implications of generative AI for the relationships between actors in the evaluations system. This perspective was considered as particularly relevant because it looks beyond the technical implications and considers the consequences of technology on existing ecosystems like the evaluation system. In this environment, diverse actors collaborate in various capacities, often following practices refined over decades, with the overarching goal of ensuring the meaningful and effective utilization of taxpayer funds.

In our discussion of generative AI, we assumed that it is already in use and that at least on an informal level new practices are already emerging. As a result, we considered it necessary to deal with questions concerning the preservation of the basic conditions of a functioning evaluation system (e.g. with regard to establishing trust - both in the results and towards the users/actors). Therefore, the focus was less on ethical or methodological issues, connected with the use of generative AI, than on the consequences on a systemic level by using this technology. In such a system we have to assume that the use of generative AI by singular or all actors will have an impact on each other's relationships – even if one or more actors are restrictive or hesitant in the use. Consequently, the development and its ensuing consequences can evolve into a self-perpetuating cycle, transcending the control of individual actors.

Therefore, we present a discussion paper resulting from the exchange and the contributions in the working group. This paper consists of initial considerations regarding the potentially changing relationships between actors in the evaluation system, triggered by the emergence of AI. It offers a basis for discussion between actors in the evaluation system who are currently confronted with the need to consider the significance and use of this technology.

The leading question was: How will relationships between actors change in an evaluation system when this system is confronted with generative artificial intelligence (AI)? The

Evaluation Systems and the Implications of AI: We understand the evaluation system as a well-established network with known actors, more or less codified standards, and the possibility for exchange and relationship maintenance. As actors, we consider those commissioning evaluations, institutions who carry them out, individual evaluators, and the

owners of the respective evaluation object. The sum of these actors and their activities form the evaluation system. This evaluation system must perform in at least four for our perspective relevant dimensions in order to fulfill its purpose:

- Dimension 1: High demand for trust. High expectations for trustworthy handling of data and evidence.
- Dimension 2: Reliability of knowledge production and reproducibility.
- Dimension 3: Speed and (cost) efficiency of evidence production for decision-making.
- Dimension 4: Specific need for contextualization and sector knowledge that has traditionally been considered a "black box" within the evaluation system

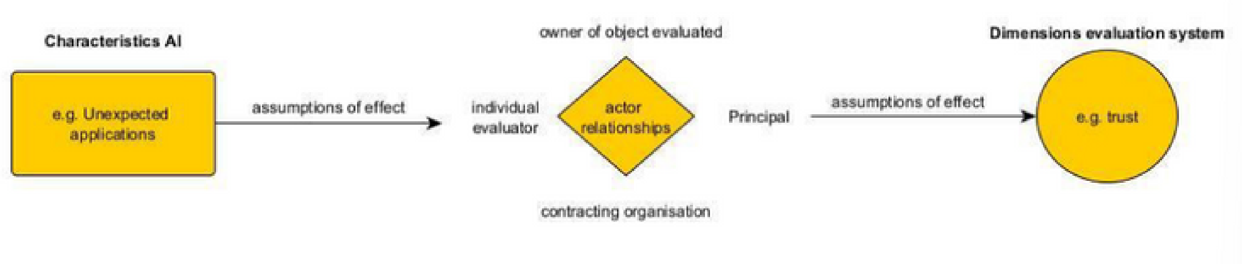
In the paper, we describe sets of positions the actors might have regarding AI and how these may affect to the functionality of the evaluation system based on the four Dimensions. Against this background, we understood AI as an “actant” in the evaluation system. This means that, unlike human actors, AI cannot act independently, but can still play an active role in a relationship between actors. The characteristics of an AI may therefore influence these relationships in a system and thereby the system as such. Based on the work of Dell'Acqua et al (2023) we found three characteristics of AI to be particularly relevant for the model and included them in it:

- Unexpected applications: meaning that nobody can predict how individual users will deploy generative AI for evaluation and where the boundaries of the application lie
- Direct enhancement of individual performance: because no large infrastructure investments are necessary, everybody can make use of generative AI and it is not clear where and how potential efficiency gains occur and who uses it for which task.
- Relative opacity: This feature of AI describes the fact, that it is difficult to assess, whether a result is viable or whether it is just a good invention of the model. This is what Dell’Acqua et al identify as a “jagged frontier”. The “jagged frontier” refers to the uneven and unpredictable boundary that makes it different to assess whether generative AI can handle a task effectively, or whether it is just producing problems. While generative AI can perform some complex tasks surprisingly well, it may struggle with simpler tasks.

The thought model is thus composed of four actors, three characteristics of AI and four Dimensions of the evaluation system, which can now be weighed up against each other in a structured manner.

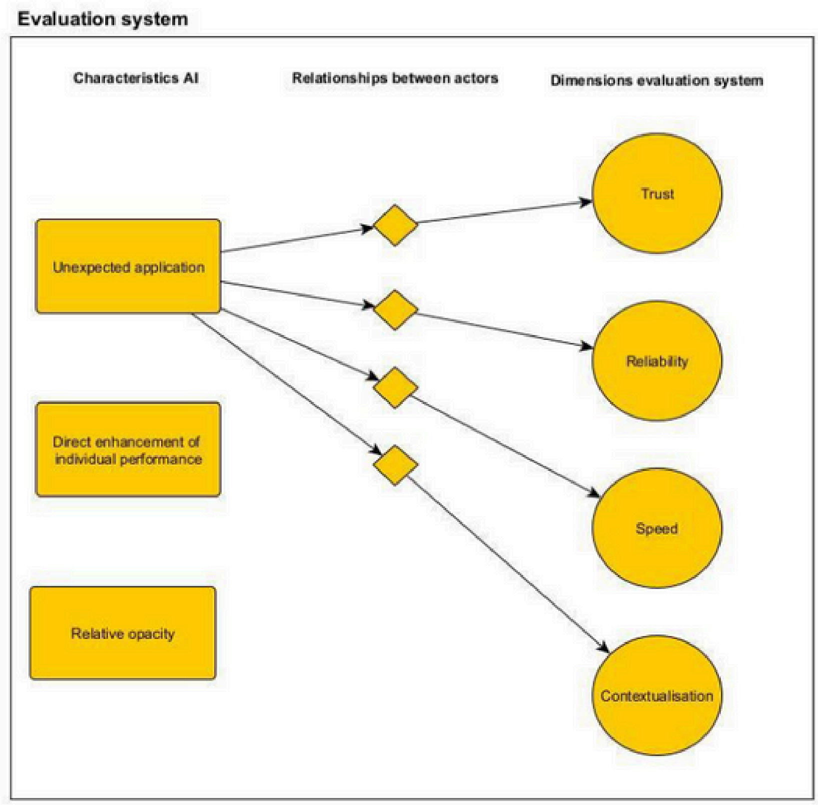
The Modell: Our model should help to moderate the discussions about the implementation of AI and to focus on the most crucial, but not directly obvious consequences of its use. For this purpose, individual aspects can be selected for discussion as shown in Figure 1 and impact assumptions can be explored on this simplified basis.

FIGURE 1: SIMPLE REPRESENTATION OF A RELATIONSHIP IN THE SYSTEM



The model can be made as complex as desired by considering more or fewer actors, characteristics or dimensions at the same time. A more complex representation can be found in Figure 2.

FIGRUE 2: COMPLEX REPRESENTATION OF RELATIONSHIPS IN THE SYSTEM



In principle, the model can be adapted to the needs of the respective user, for example by exchanging actors, characteristics of AI or dimensions of the system. It enables a structured exchange within a working group or similar set-up based on defined characteristics of properties or stakeholder groups. This helps users to gain a common understanding of what the properties of AI are, for example, or which aspects of a system should be considered in particular.

Against this background, central questions are proposed that, from the working group's perspective, are worth considering as an actor in the evaluation field, and constructive adaptation possibilities to the emergence of the technology are presented for discussion. With this systemic perspective, we hope to contribute another viewpoint to the discourse that goes beyond questions of data protection, ethics, and scientific reliability in the use of AI. We would be very pleased to share this contribution with the fteval community and expand upon their perspective.

REFERENCES:

- Dell'Acqua, Fabrizio and McFowland, Edward and Mollick, Ethan R. and Lifshitz-Assaf, Hila and Kellogg, Katherine and Rajendran, Saran and Krayer, Lisa and Candelon, FranÁois and Lakhani, Karim R., Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality (September 15, 2023). Harvard Business School Technology & Operations Mgt. Unit Working Paper No. 24-013, Available at SSRN: https://ssrn.com/abstract=4573321 or http://dx.doi.org/10.2139/ssrn.4573321

Jonathan Dudek, Jeroen van Honk, Isabel Basson, Carole de Bordes, Ismael Rafols, Rodrigo Costas
CWTS, Leiden University, Netherlands
CWTS, Leiden University, Netherlands
CREST, Stellenbosch University, South Africa
CWTS, Leiden University, Netherlands
CWTS, Leiden University, Netherlands
CWTS, Leiden University, Netherlands

FUNDING FLOWS IN AFRICA FROM FOREIGN SOURCES: TOPICS, DEPENDENCIES AND IMPLICATIONS

ABSTRACT

The research systems of African countries have a significant dependence on international collaboration and funding, which is likely to affect their research agendas. This dependency is particularly relevant because large international funders include philanthropic foundations whose priorities do not necessarily respond to local public priorities or global scientific communities (Vessuri, 2017).

The analysis of research funding is an emerging topic in science and technology studies, both due to the growing interest in understanding the effects of specific instruments in science (Lepori et al., 2023) and due to the desire to prioritize research towards social needs and aspirations (Ciarli & R ‡ fols, 2019). The interest in understanding the influence of investments in science towards well-being is particularly important in developing societies with large social inequalities.

This study maps the volume of funding flows in Africa to countries, organizations, scientific fields and topics, in order to understand their relative importance and influence. In the abstract, we present analyses based only on the acknowledgments mentions of funders, but in the conference we will also have grant data, based on the Dimensions database (Herzog et al., 2020). In a past study (Kozma et al., 2018), a comprehensive funding landscape was provided for all African countries using the Web of Science (Science Citation Index Expanded - SCIE) as a database for the period 2019-2014.

The following research questions are addressed in this work: What is the number and share of African publications reporting funding? Are there differences across African countries? Who are the main funders in Africa? How do their funded publications behave in terms of collaboration patterns? What is the overall disciplinary presence of funding and funders in the continent? Are top funders specialized in certain disciplines?

The Dimensions database was used in this study for capturing funding acknowledgements and grant data. The analysis period goes from 2010 up to and including 2021. Given the descriptive nature of this analysis and the fact that no citation analysis is performed in this

study, we have included all document types captured by Dimensions. The geographical spectrum of the countries analysed includes the 58 African countries (continent='AF'), identified by their ISO-3166 alpha2 code that appears in the Geonames database (https://www.geonames.org/countries/). The disciplinary perspective is established using the five main research fields used in the Leiden Ranking classification (https://www.leidenranking.com/information/fields).

The north-south divide in Africa regarding the funding reported in scientific publications is observed again. Large North-African countries like Egypt, Nigeria, Tunisia, Morocco, Algeria and Ethiopia present substantially lower shares of publications with funding information, usually below 20% of all publications. A similar result was also observed for some of these countries by El-Ouahi (2023), using the Web of Science database. According to El-Ouahi, these low levels of funding sources mentioned in publications can be associated with block funding structures, in which lecturers or researchers are funded by universities or their governments without necessarily reporting them as funders in their publications. This is an aspect deserving more attention in the further development of the project.

Table 1 presents the top15 main funders in Africa by the number of publications in the continent reporting them as funders. Column ‘Pubs’ captures the total number of publications that report funding from each of the funders (for reasons of simplicity, full counting is applied, so if a publication has more than one funder, it is counted for each funder). Column ‘% collab’ captures the percentage of publications of each funder that involves some degree of collaboration (i.e., having authors from more than one affiliation), while column ‘% int collab’ shows the percentage of publications involving international collaboration (i.e., having authors from more than one country).

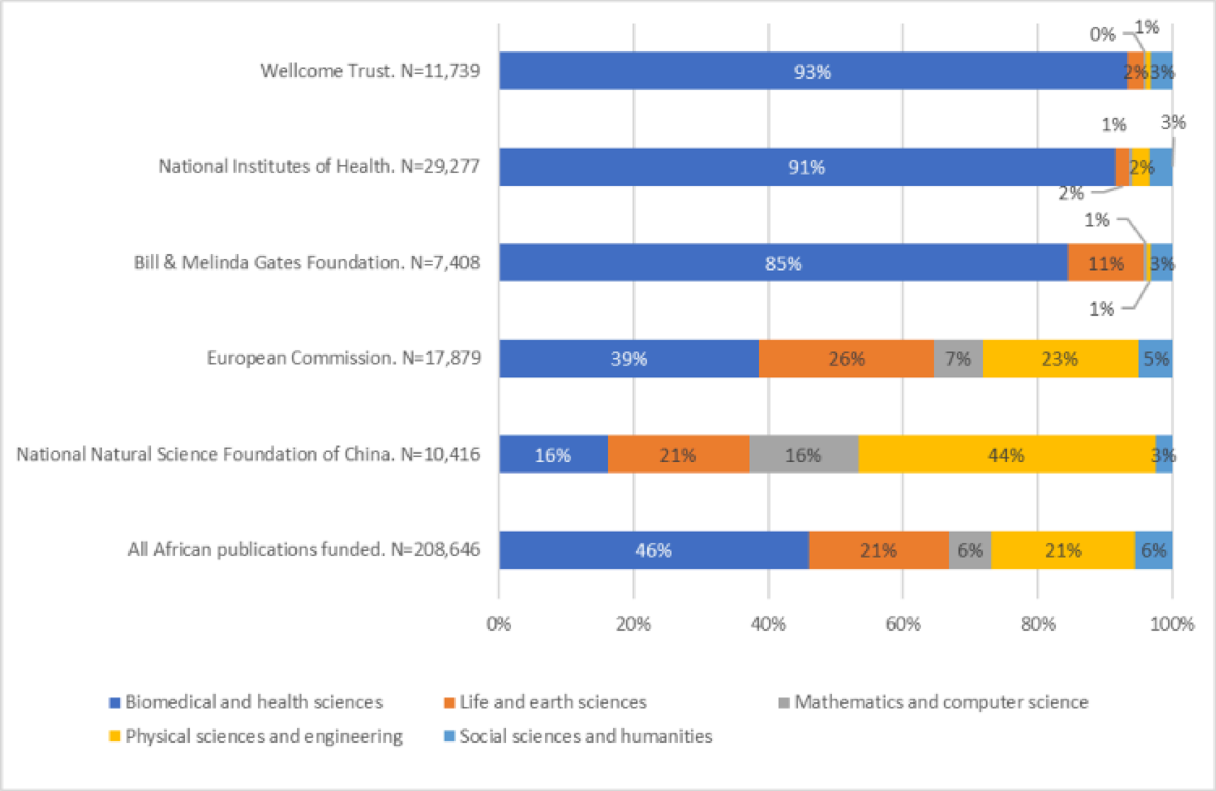
TABLE 1. TOP 15 FUNDERS IN AFRICA IN DIMENSIONS ACCORDING TO ACKNOWLEDGEMENTS (2010-2021).

Funder	Country	Continent	Pubs	% collab	% int
National Research Foundation (NRF)	***South Africa***	***Africa***	***36429***	***60.50%***	***43.49%***
National Institutes of Health (NIH)	USA	North America	29277	95.34%	92.23%
European Commission (EC)	European Union	Europe	17879	94.93%	92.77%
Wellcome Trust	United Kingdom	Europe	11739	94.94%	90.60%
National Natural Science Foundation of China (NSFC)	China	Asia	10416	99.54%	99.38%
Medical Research Council (MRC)	United Kingdom	Europe	9871	95.83%	93.18%
Department of Science and Technology (DST)	***South Africa***	***Africa***	***7681***	***66.27%***	***47.60%***
Bill & Melinda Gates Foundation	USA	North America	7408	96.13%	93.35%
Japan Society for the Promotion of Science (JSPS)	Japan	Asia	6400	99.36%	99.25%
Deutsche Forschungsgemeinschaft (DFG)	Germany	Europe	6288	98.30%	97.63%
South African Medical Research Council (SAMRC)	***South Africa***	***Africa***	***5824***	***76.41%***	***55.77%***
World Health Organization (WHO)	Switzerland	Europe	5408	92.29%	86.76%
French National Centre for Scientific Research (CNRS)	France	Europe	5105	97.71%	97.02%
Chinese Academy of Sciences (CAS)	China	Asia	2907	99.38%	98.83%
Dutch Research Council (NWO)	Netherlands	Europe	2659	98,16%	96,65%

The largest share of African publications reporting funding is shown for Life and earth sciences (34%), followed by the Biomedical and health sciences (28%) and the Physical sciences and engineering (26%). Publications from the Social sciences and humanities (11%) as well as Mathematics and computer science (11%) feature less funding information. This suggests that these two fields are relatively underfunded in Africa.

Figure 1 shows the disciplinary profile of the publications funded by the five main non-African funders on the continent. The first observation based is that African-funded publications overall have a strong concentration in the Biomedical & health sciences, with almost half of all funded publications classified in this field. This is followed by the Life & earth sciences and Physical sciences & engineering. Mathematics & computer sciences and Social sciences & humanities both take a small share.

FIGURE 1: SHARE OF PUBLICATIONS FUNDED BY A SELECTION OF MAIN FUNDERS' MAIN FIELDS OF SCIENCE.



The Wellcome Trust, the National Institutes of Health and the Bill & Melinda Gates Foundation have a strong focus on Biomedical & health sciences, with the latter also exhibiting some degree of focus on the Life & earth sciences. The European Commission and the National Natural Science Foundation of China both present a more multidisciplinary distribution, although the European Commission has a relatively stronger focus on Life & earth sciences, while the Chinese funding organization has a stronger orientation towards the Physical sciences & engineering

Main findings and conclusions: This exploratory study presents funding flows in Africa according to publication acknowledgements using Dimensions data. During the conference will plan to present an analysis of funded projects based on more detailed grant data, including funding amounts, grantees, etc., which is the most innovative parts of the overall study.

International funders have a strong presence in the African continent. Funders like the European Commission, the NIH, the Wellcome Trust, the Chinese National Natural Science Foundation, or the Bill & Melinda Gates Foundation dominate the funding landscape. Also, publications funded by these international organizations are usually associated with internationally collaborative publications, suggesting a collaborative linkage between researchers from the countries of the funders and African researchers. Additionally, the disciplinary analysis shows a stronger presence of funding information reported in publications from the disciplines of Biomedical & health sciences, and Life and earth sciences, while the production in Physical sciences & engineering shows much lower shares of funding as compared with what is observed globally.

Our results also show the differentiated topical interests of international funders, with some of the main funders (NIH, Wellcome Trust and Bill and Melinda Gates Foundation) heavily focusing on biomedical research, while the European Commission or the National Natural Science Foundation of China having a stronger focus on life and earth sciences, and physics and engineering respectively. Such differentiated patterns open the question of their role in setting the disciplinary research agenda in the continent and may suggest a need for more coordinated strategies, in order to find a balance, e.g., by considering relatively under-funded fields like social sciences and humanities, or mathematics and computer sciences.

The policy implications of this dependency on international funders for agenda-setting according to national and local needs will be discussed in the conference (Vessuri, 2017).

This work was carried out with the aid of a grant from the International Development Research Centre (IDRC), Ottawa, Canada. The views expressed herein do not necessarily represent those of IDRC or its Board of Governors.

REFERENCES:

- Ciarli, T., & R ‡ fols, I. (2019). The relation between research priorities and societal demands: The case of rice. Research Policy, 48(4), Article 4. https://doi.org/10.1016/j.respol.2018.10.027
- Herzog, C., Hook, D., & Konkiel, S. (2020). Dimensions: Bringing down barriers between scientometricians and data. Quantitative Science Studies, 1(1), 387–395.

- Kozma, C., Calero Medina, C., & Costas, R. (2018). Research funding landscapes in Africa. The next Generation of Scientists in Africa, 26–42. Lepori, B., Jongbloed, B., & Hicks, D. (2023). Introduction to the Handbook of Public Funding of Research: Understanding vertical and horizontal complexities. In B.
- Lepori, B. Jongbloed, & D. Hicks (Eds.), Handbook of Public Funding of Research (pp. 1–19). Edward Elgar Publishing. https://doi.org/10.4337/9781800883086.00005
- Vessuri, H. (2017). From Science as "Development Assistance" to "Global Philanthropy". In D. Tyfield, R. Lave, S. Randalls, & C. Thorpe (Eds.), The Routledge Handbook of the Political Economy of Science (1st ed., pp. 405–415). Routledge. https://doi.org/10.4324/9781315685397- 36

EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

This publication is based upon work from COST Action CA20112 PROFEEDBACK, supported by COST (European Cooperation in Science and Technology). COST (European Cooperation in Science and Technology) is a funding agency for research and innovation networks. Our Actions help connect research initiatives across Europe and enable scientists to grow their ideas by sharing them with their peers. This boosts their research, career and innovation. Visit www.cost.eu

WWW.COST.EU

Funded by the European Union

FOLLOW US!

WWW.PROFEEDBACK.EU