53

fteval JOURNAL for Research and Technology Evaluation

ANETE VINGRE, PETER KOLARZ AND BILLY BRYAN

ON YOUR MARKS, GET SET, FUND!
RAPID RESPONSES TO THE COVID-19 PANDEMIC

FLORIAN WITTMANN, FLORIAN ROTH, MIRIAM HUFNAGL, RALF LINDNER AND MERVE YORULMAZ

TOWARDS A FRAMEWORK FOR IMPACT ASSESSMENT FOR MISSION-ORIENTED INNOVATION POLICIES.
A FORMATIVE TOOLBOX APPROACH

LISE MOAWAD AND CORNELIA SCHENDZIELORZ

METAMORPHOSES AND PERFORMATIVITY.
TRANSFORMATIVE R&I POLICIES AND THE NORM(ALIS)ING EFFECT OF SOCIETAL IMPACT

ISSUE 53 | APRIL 2022

FEDERAL MINISTRY OF EDUCATION, SCIENCE AND RESEARCH
Minoritenplatz 5, 1014 Vienna
Mag.ª Irene Danler
E: irene.danler@bmbwf.gv.at
Mag.ª Simone Mesner
E: simone.mesner@bmbwf.gv.at

FEDERAL MINISTRY FOR DIGITAL AND ECONOMIC AFFAIRS
Stubenring 1, 1014 Vienna
Mag.ª Sabine Pohoryles-Drexel
E: sabine.pohoryles-drexel@ bmdw.gv.at

FEDERAL MINISTRY FOR CLIMATE ACTION, ENVIRONMENT, ENERGY, MOBILITY, INNOVATION AND TECHNOLOGY
Radetzkystraße 2, 1030 Vienna
Dr. Rupert Pichler
E: rupert.pichler@bmk.gv.at
MSc (WU) Philipp Witibschlager, BA
E: philipp.witibschlager@bmk.gv.at

ACR – AUSTRIAN COOPERATIVE RESEARCH
Sensengasse 1, 1010 Vienna
Dr.in Sonja Sheikh
E: sheikh@acr.ac.at

AUSTRIAN COUNCIL FOR RESEARCH AND TECHNOLOGY DEVELOPMENT
Pestalozzigasse 4/DG 1, 1010 Vienna
Sascha Ruhland, M.A.
E: s.ruhland@rat-fte.at

AIT – AUSTRIAN INSTITUTE OF TECHNOLOGY
Giefinggasse 4, 1210 Vienna
Mag. Michael Dinges
E: michael.dinges@ait.ac.at
Anna Wang
E: anna.wang@ait.ac.at

AQ AUSTRIA – AGENCY FOR QUALITY ASSURANCE AND ACCREDITATION AUSTRIA
Franz-Klein-Gasse 5, 1190 Vienna
Dr.in Elisabeth Froschauer-Neuhauser
E: elisabeth.froschauer-neuhauser@aq.ac.at
Dr.in Eva Maria Freiberger
E: eva.maria.freiberger@aq.ac.at

AWS – AUSTRIA WIRTSCHAFTSSERVICE GMBH
Walcherstraße 11A, 1020 Vienna
Dr.in Elvira Welzig
E: e.welzig@aws.at
Mag. Norbert Knoll
E: n.knoll@awsg.at

CDG – CHRISTIAN DOPPLER RESEARCH ASSOCIATION
Boltzmanngasse 20, 1090 Vienna
DIin Magª. Brigitte Müller
E: brigitte.mueller@cdg.ac.at

CONVELOP – COOPERATIVE KNOWLEDGE DESIGN GMBH
Bürgergasse 8-10/I, 8010 Graz
DIin Drin. Karin Grasenick
E: karin.grasenick@convelop.at
Erdbergstraße 82/4, 1030 Wien
Konstantin Melidis
E: konstantin.melidis@convelop.at

FFG – AUSTRIAN RESEARCH PROMOTION AGENCY
Haus der Forschung,
Sensengasse 1, 1090 Vienna
DIin Dr.in Sabine Mayer
E: sabine.mayer@ffg.at
Mag. Leonhard Jörg
E: leonhard.joerg@ffg.at

FWF – AUSTRIAN SCIENCE FUND
Haus der Forschung
Sensengasse 1, 1090 Vienna
Dr. Falk Reckling
E: falk.reckling@fwf.ac.at

IHS – INSTITUTE FOR ADVANCED STUDIES
Josefstädter Straße 39, 1080 Vienna
Dr.in Angela Wroblewski
E: wroblews@ihs.ac.at

INDUSTRIEWISSENSCHAFTLICHES INSTITUT – IWI
Mittersteig 10, 1050 Wien
FH-Hon.Prof. Dr. Dr. Herwig W. Schneider
E: schneider@iwi.ac.at
Mag. Philipp Brunner
E: brunner@iwi.ac.at

JOANNEUM RESEARCH FORSCHUNGSGESELLSCHAFT MBH
Haus der Forschung
Sensengasse 1, 1090 Vienna
Mag. Wolfgang Polt
E: wolfgang.polt@joanneum.at
Mag. Jürgen Streicher
E: juergen.streicher@joanneum.at

KMU FORSCHUNG AUSTRIA – AUSTRIAN INSTITUTE FOR SME RESEARCH
Gusshausstraße 8, 1040 Vienna
Mag. Peter Kaufmann
E: p.kaufmann@kmuforschung.ac.at
Dr. Mario Steyer
E: m.steyer@kmuforschung.ac.at

LUDWIG BOLTZMANN GESELLSCHAFT
Nußdorfer Straße 64, 1090 Vienna
Mag. Patrick Lehner
E: patrick.lehner@lbg.ac.at

OEAD - AGENCY FOR EDUCATION AND INTERNATIONALISATION
Ebendorferstraße 7, 1010 Wien
Mag. Ulrich Hörmann
E: ulrich.hoermann@oead.at

OEAW – AUSTRIAN ACADEMY OF SCIENCE
Dr. Ignaz Seipel-Platz 2, 1010 Vienna
Nikolaus Göth, MSc
E: nikolaus.goeth@oeaw.ac.at

TECHNOLPOLIS GROUP AUSTRIA
Rudolfsplatz 12/11, 1010 Vienna
Mag.ª Katharina Warta
E: warta@technopolis-group.com
Tobias Dudenbostel, MA
E: tobias.dudenbostel@technopolis-group.com

VIENNA BUSINESS AGENCY. A SERVICE OFFERED BY THE CITY OF VIENNA.
Mariahilfer Straße 20, 1070 Vienna
Robert Mayer-Unterholzer
E: mayer-unterholzner@wirtschaftsagentur.at
DI Mag. Alfried Braumann
E: braumann@wirtschaftsagentur.at

WIFO – AUSTRIAN INSTITUTE OF ECONOMIC RESEARCH
Arsenal, Objekt 20
Postfach 91, 1103 Vienna
Dr. Jürgen Janger
E: juergen.janger@wifo.ac.at
Dr. Andreas Reinstaller
E: andreas.reinstaller@wifo.ac.at

WPZ RESEARCH GMBH
Mariahilfer Straße 115/16, 1060 Vienna
Dr.in Brigitte Ecker
E: brigitte.ecker@wpz-research.com

WWTF – VIENNA SCIENCE AND TECHNOLOGY FUND
Schlickgasse 3/12, 1090 Vienna
Dr. Michael Stampfer
E: michael.stampfer@wwtf.at
Dr. Michael Strassnig
E: michael.strassnig@wwtf.at

ZSI – CENTRE FOR SOCIAL INNOVATION
Linke Wienzeile 246, 1150 Vienna
Dr. Klaus Schuch
E: schuch@zsi.at
Drin Dorothea Sturn
E: sturn@zsi.at

EDITORIAL

DEAR READERS!

After its pandemic-induced postponement, the REvaluation Conference 2021|22 finally takes place 5-6 May 2022 in Vienna with more than 60 contributions in over 20 sessions. Furthermore, the conference included several panels, keynote speeches as well as the fteval Platform's 25-years anniversary enquete. We are delighted to publish with this issue of the fteval Journal for Research and Technology Policy Evaluation the Conference Proceedings, presenting 20 selected articles along the event's topical strands *Anticipation – Transformation – Resilience*, relating to some of the grand challenges of our time. We hope to inspire policy makers, programme owners, funding agencies, researchers and evaluators with interesting approaches for further developing research and innovation systems in these three domains.

EFFECTS OF THE PANDEMIC TO RESEARCH AND INNOVATION FUNDING

Rapid response to the dynamics of the COVID-19 pandemic was not only asked from the participants and the organising team of the REvaluation Conference, but also from research funders around the world. The **effects the pandemic has had on research and innovation funding** is topic of two articles: How selected funding agencies reacted in order to promptly provide financial support for research on this new global plague and what type of processes and structures proved beneficial can be read in the contribution by Vingre, Kolarz and Bryan [read Article 1 from page 8].The team of Kerlen, Kind, Rodriguez, Wangler, Zinke and Wessels focuses on business R&D and provide preliminary insights on the effects of the pandemic on innovation funding, adaptation strategies and how these can be provisionally evaluated. In monitoring and evaluation processes running since summer 2020, five German innovation funding programmes were analysed - their first results are presented in this article. In addition, the authors propose an impact model that takes this external shock on ongoing innovation projects into account.

Most evaluations build on a programme theory. Consequently, an impact model interweaving internal with external factors usually forms the starting point of a programme evaluation. To enable an appropriate consideration of the effect of the pandemic, the authors developed an impact model to be applied in a generalised form to different funding measures in the field of research, technology and innovation policy [read Article 2 from page 18].

WHAT DOES TRANSFORMATION MEAN FOR POLICY EVALUATION?

Anticipation and resilience in research and innovation policy-making were but two of the topical strands of this year's REvaluation Conference. The largest strand Transformation puts a spotlight on the role of science and science policy for societal development towards social, environmental, cultural and economic sustainability. Questions on how to evaluate the effects of transformative policies - while the system under consideration itself is transforming - remain intricate and heavily discussed.

For these Conference Proceedings we collected eleven contributions on transformation issues in research policy evaluation. Three articles deal with **conceptual frameworks for the evaluation of mission-oriented and transformative policies**: in Article 3 [read from page 31], Wittmann, Roth, Hufnagl, Lindner and Yorulmaz share their considerations for a comprehensive toolbox for impact assessment of missions and the engagement of the many different perspectives relevant for the success of a mission. While providing a handy guidance for evaluation, the authors also showcase the caveats that an active inclusion of a large number of stakeholders has.

At the level of process indicators, Schuck-Zöller and her colleagues suggest a formative evaluation scheme for climate services. From the experience of their work at the Helmholtz Institute for Climate Service Science, where interaction between stakeholders is their daily business, we can learn how success criteria can be implemented in co-creative research and processes [read Article 4 from page 43].

Dinges, Kerlen, Kaufmann, Wang, Toepel, Kofler, Meyer and Wieser take a more global perspective in their discussion of theories of change for transformation-oriented policies. Using the example of the German 7th Energy Research Programme, they contribute innovation system thinking to the analysis of transformative environments [read Article 5 from page 57].

Two contributions illustrate what transformative research policies can stand for at the **regional innovation** level: in their Article Mena Jara, Meijer, Heimericks and Willemse discuss how smart specialisation and **responsible research and innovation (RRI)** can be integrated in territorial stakeholder networks and policy systems [read Article 6 from page 69]. Sourcing from several projects treating the implications of

monitoring and evaluation ecosystem, Tjitske and her colleagues offer first considerations to theorise their concept of evaluative conversation as an approach for improving regional RRI [read Article 7 from page 77].

The two articles that follow build a nice transition as they also contribute to the **RRI discourse by offering conceptual impact frameworks**: Moawad and Schendzielorz observe how transformative research and innovation (R&I) policies have a norming and normalising effect on what is considered as societal impact [read Article 8 from page 85]. In his article, Brasil illustrates the application of a multidimensional self-evaluation approach in the Brazilian graduate system and reflects its potential for improvement [read Article 9 from page 97].

How concepts that assess impact of RRI programmes can be **co-created** is a special focus of another two contributions: Wailzer and Soyer present their co-developed impact model for evaluating the societal impact of participatory research approaches used at the Open Innovation in Science Center at the Ludwig Boltzmann Society [read Article 10 from page 107]. Yorulmaz and Bührer co-created a template to record and structure the potential impact arising from RRI, taking into consideration the scientific, societal as well as economic impacts [read Article 11 from page 118].

Examples for assessments of specific transformative policies are highlighted in the article by Schneuwly and Chandler on European agri-food R&I investments [read Article 12 from page 126], and Seus and Stadler monitored a city lab process in Mannheim, tracing what aspects of social innovation occurred in the energy sector [read Article 13 from page 133].

METHODOLOGICAL ADVANCES IN RESEARCH AND INNOVATION POLICY EVALUATION

The final section presents insights gained in specific evaluation endeavours. In our joint conference series, IFRIS, Fraunhofer ISI and fteval, traditionally give stage to discussing **methodological advances** in research and innovation policy evaluation, and research on science policy in general. The following seven articles focus on the multifaceted lessons learned and reflections undertaken: in five articles, the authors present **specific examples of R&I programmes** and their evaluation as well as the **implementing Agencies' perspectives** on these programmes. The last two articles present general methodological reflections.

Kaisler and Palfinger contribute with a presentation of how funding, facilitating and evaluating participatory research approaches in Austria can be realised by using the example of the Ludwig Boltzmann Society's programme on Patient and Public Involvement and Engagement. [read Article 14 from page 141]. Dudenbostel draws lessons from the evaluation of the Philipp Schwartz Initiative, a German programme that supports researchers, who had to flee from their home countries, in establishing their scientific career in German universities [read Article 15 from page 147]. Miyajima, Isshiki, Kunugi and Uesaka asked whether successful market introduction of products and services, resulting from funded projects, can be predicted using ongoing R&D evaluation data of the Japanese Agency NEDO. [read Article 16 from page 153]. For the Austrian Research Promotion Agency FFG, Landon and Hochreiter introduce a randomised controlled trial approach and reflect how this and other experimental approaches have been established in the FFG [read Article 17 from page 160]. Finally, the ways and means with which Bpifrance seeks to understand the impact of SME's innovation projects is described by Brun [read Article 18 from page 169].

Taking the lessons learned to a more general level, further **methodological reflections** are raised by the last two articles. Rothgang and Lageman aim at understanding why approaches of systems analysis so far have only played a small role in innovation policy evaluation. They make suggestions how the detected epistemological and institutional obstacles can be overcome [read Article 19 from page 181]. And finally, Uhrig and Spanó reflect on how the model of "key impact pathways" can serve on strategic, institutional and on project level [read Article 20 from page 192].

Holding our journal issue of 204 pages in hands is the result of many hours of writing, reading, reworking, formatting, reviewing, revising, proofreading, feedback looping, mailing back and forth, collecting images and bits and pieces, checking, coordinating and great cooperation amongst the editorial team. We thank all contributors for the efforts and the creativity that went into this issue of the fteval Journal for Research and Technology Policy Evaluation!

We wish you a good read and interesting insights!

Klaus Schuch, Elisabeth Nindl & Isabella Wagner

April 2022

TABLE OF CONTENTS

ON YOUR MARKS, GET SET, FUND! RAPID RESPONSES TO THE COVID-19 PANDEMIC

ANETE VINGRE, PETER KOLARZ AND BILLY BRYAN

DOI: 10.22163/fteval.2022.538

ABSTRACT

This paper presents findings from an analysis of seven multidisciplinary national research funders' responses to COVID-19. We posit that while some parts of research and innovation funding responses to COVID-19 were 'pandemic responses' in the conventional biomedical sense, other parts were thematically far broader and are better termed 'societal emergency' funding. This type of funding activity was unprecedented for many funders. Yet, it may signal a new/additional mission for research funders, which may be required to tackle future societal emergencies, medical or otherwise. Urgency (i.e., the need to deploy funding quickly) is a key distinguishing theme in these funding activities. This paper explores the different techniques that funders used to substantially speed up their application and assessment processes to ensure research on COVID-19 could commence as quickly as possible. Funders used a range of approaches, both before application submission (call design, application lengths and formats) and after (review and decision-making processes). Our research highlights a series of trade-offs, at the heart of which are concerns around simultaneously ensuring the required speed as well as the quality of funding-decisions. We extract some recommendations for what a generic 'societal emergency' funding toolkit might include to optimally manage these tensions in case national research funders are called upon again to respond to future crises.

INTRODUCTION

This paper presents findings from an analysis of seven multidisciplinary national research funders' responses to COVID-19. The background work for this study was conducted as part of a recently completed process review of UKRI's research and innovation (R&I) response to COVID-19 (Kolarz et al, 2021), which included a substantial international comparative dimension.

We begin by positing that funders' responses went decisively beyond pandemic response in the conventional sense and amounted to a largely unprecedented type of R&I funding. COVID-19 marked the first time many research funders were called on to rapidly mobilise researchers from a broad range of disciplines and fund large bodies of research as rapidly as possible to respond to a major unfolding societal crisis. And it may not be the last.

OECD figures help to give a sense of the scale of the global R&I funding response to COVID-19: the OECD Science, Technology, and Innovation Policy Compass COVID-19 tracker database lists 702 policy initiatives targeting COVID-19 pandemic across the OECD countries (OECD, 2021b). The OECD estimates that over $7b were unlocked in the first nine months of 2020 (OECD, 2021a). Funders introduced new measures and extended the duration of ongoing research and deadlines for new calls for applications (Stoye, 2020). In some cases, funders made only one-off investments in 2020, whereas other measures continue to operate in 2021.

Across many of these measures, the need for a rapid response meant that 'business-as-usual' funding processes had to change – at least by accelerating existing processes, or indeed by modifying them more substantively. At the same time, funders had to ensure that acceleration and modification did not compromise the quality of decision-making, i.e., that they still funded high-quality research.

Studying and understanding these funding responses has merit from an evaluative point of view: did funders perform as well as they could have done? But COVID-19, its countermeasures and consequences also highlighted the critical role that public research funders play in major societal emergencies. Understood as a wider societal emergency rather than strictly as a pandemic, an assessment of funders' responses to COVID-19 can therefore also help to define some parameters for a rapid-response toolkit suited to future crises, health-related or otherwise.

Academic literature on the need for rapid research and the response of funders is scarce. It focuses either on the implications for research practice (see, for example, Richardson et al, 2021; Lurie et al, 2021) or on funding for clinical research. For example, the main recommendation of Sigfrid et al.'s 2020 review of the academic literature on clinical research responses to pandemics (including COVID-19) was to increase STI preparedness before a pandemic rather than a purely reactive response. Dedicated emergency funding for the rapid release of funds, strong international collaborations and community engagement (e.g., involving affected communities in programme design) from the outset were cited as key enablers for a successful rapid STI response.

In its most recent STI Outlook, the OECD has roughly outlined some of the approaches taken by funders and highlighted some of the challenges (OECD, 2021b). However, we are not aware of published work exploring this topic in detail with primary data from funders or looking across multidisciplinary research responses.

APPROACH

We used an exploratory comparative case study approach for this study to allow for an iterative analysis of each case (funder) with a final comparison of emergent themes and explanations (Mills et al. 2010). This approach is particularly useful for analysing organisational processes and change in response to a common external problem (i.e., the research needs to understand and address the impact of COVID-19).

The study covers seven research funders: the Dutch Research Council NWO and its sister organisation for health research ZonMw, the German Research Council DFG, the UK national funding agency UKRI, National Research Council of Canada (NRC), the Japan Science and Technology Agency (JST), the Ministry of Science and Technology of Taiwan (MoST) and the National Science Foundation (NSF) in the USA.

The study used desk research and interviews. The information available on the funders' websites, meeting protocols, reports and grey literature were used to conduct the desk research. Semi-structured interviews with representatives of funders[1] were conducted to fact-check the findings of desk research and gain insights into challenges the funders faced in introducing and implementing their responses to COVID-19. Interviews were conducted with personnel directly involved in the design and delivery of the response and had strategic oversight of their funder's mission and role in the respective R&I system.

A common template was used across cases for structured data collection. The analysis was performed from April to July 2021, when most funders had completed the first rounds of response mechanisms and were ready to reflect on the first lessons learned.

This study has some limitations. First, it is based on funders' perceptions, which can be biased, though we worked to substantiate their accounts with documentary evidence where possible. Second, it covers a limited selection of funders (from developed countries) and cannot be globally representative but does provide a view of some of the most active funders in the world on this issue. Third, the success of funders' response might arguably be better judged in the longer-term post-pandemic. However, the focus of this study is around short-term response processes and learning from them for future crises. This means that funders' ability to deploy funding rapidly is a key success criterion: unlike in other funding activities, where it is usually accepted that impacts may only appear 'downstream', often after many years, large parts of funders' responses to COVID-19 were intended to produce impact and actionable knowledge within months. Finally, the study covers only rapid funding instruments and does not explore the effects of the pandemic or rapid response on other funding priorities and instruments. Although such a wholistic understanding would be very relevant, it merits a separate investigation and is an area for further research.

FROM 'PANDEMIC RESPONSE' TO 'SOCIETAL EMERGENCY'

The mission of each funder was not always discernible or fully predefined, but urgency is a core theme that characterises all funders' responses. Accordingly, all funders introduced programmes aiming to quickly support research to understand and address the impacts of COVID-19. Table 1 provides a brief overview of the reviewed funders' responses to the COVID-19 pandemic.

Table 1 Reviewed funders' response to COVID-19

Funder	Response to COVID-19	Overall funding and number of supported awards
NWO and ZonMw, Netherlands	Corona: fast-track data call for applications (NWO) Two waves of the research programme COVID-19 (NWO and ZonMw) 'Virus Outbreak Data Access Network' initiative for data sharing	COVID-19 programme provided €56.5m to 235 awards. Fast-track data programme provided €1.5m to 34 awards.
DFG, Germany	Set up of Interdisciplinary Commission for Pandemic Research to steer the response to COVID-19 Call for Multidisciplinary Research into Epidemics and Pandemics in response to the Outbreak of SARS-CoV-2 COVID-19 Focus Funding instrument	COVID-19 Focus Funding provided €3.6m to 33 awards (first call). Call for multidisciplinary research into epidemics and pandemics provided €30m to 50 awards.
UKRI, United Kingdom	Fifteen COVID-19 research programmes and interventions, including: joint agency programmes (e.g., UKRI-NIHR programme funded the Oxford/AZ Vaccine), international programmes (e.g., GECO - Global Effort on COVID-19 Health Research), UKRI COVID-19 Agile R&I response call, infrastructure (Vaccine Manufacturing Innovation Centre [VMIC]), and policy programmes (National Core Studies)	UKRI provided £647.2m to 1,057 awards from February 2020 to May 2021. Significant investments were the VMIC (£200.2m) and the UKRI Agile Open call (£172.5m)
National Research Council Canada (NRC)	The Pandemic Response Challenge programme and new Vice president to lead the programme New infrastructure projects (manufacturing, clinical trial centres) Industrial Research Assistance Programme challenges for businesses	Pandemic Response Challenge programme provided €15m to 6 awards in 2020.

1 Except NSF, who could not be reached for consultation.

Funder	Response to COVID-19	Overall funding and number of supported awards
Ministry of Science and Technology of Taiwan (MoST)	MoST introduced new accelerated funding instrument with supplementary funding to support: Short-term missions focusing on quick solutions for testing, treatment, vaccines Long-term missions focusing on epidemiology and policy making	C-19 research call provided €30m.
Japan Science and Technology Agency (JST)	Covid call in the Strategic Basic Research Programme J-RAPID programme funding international collaborative research	J-RAPID provided €4.1m to 11 awards. CREST provided €30m to 10 awards in 2020.
National Science foundation (NSF), USA	NSF responded to COVID-19 by investing $75m in fast response research through its RAPID mechanism previously deployed to respond to other emergencies.	RAPID provided $75m to over 1000 awards up to end of October 2020.

Among the earliest and most evident observations in our research was that there is an important core distinction within the R&I responses to COVID-19, namely between what we term 'pandemic response' conventionally defined, and a much broader element we term 'societal emergency' funding. The remainder of this paper focusses on the latter, so we describe the distinction here.

While the scale of the COVID-19 pandemic was unprecedented, little is new about research funders conducting or aiding a pandemic response. Research funders in the biomedical sciences have coordinated efforts and provided rapid funding before, responding to other health emergencies such as Zika and Ebola outbreaks (Oliveira et al, 2020). Besides such past experiences, research funders conducting a pandemic response to COVID-19 could also rely on guidance and standard processes from Glopid-R and the WHO R&D Blueprint.

Many of the funders covered in this study made substantial contributions towards tackling the pandemic itself. This focused on therapeutics, diagnostics, and understanding the spread of the disease, typically in a biological/genetic sense, though occasionally also involving social scientific work. Often, these pandemic response activities were centred on a small number of known competence centres within a few specific fields.

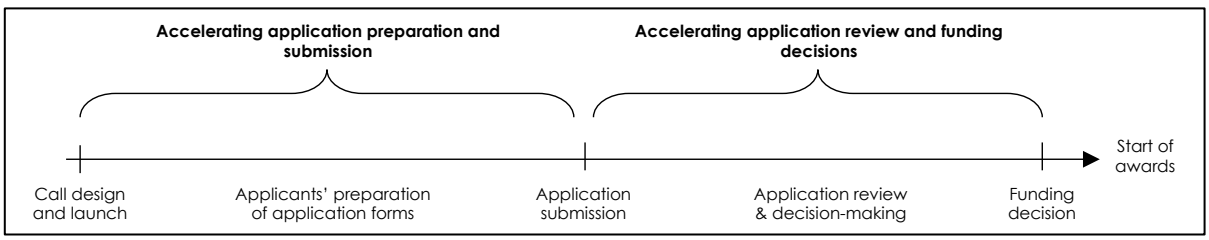
Beyond the early months of 2020, the R&I funding responses expanded beyond what is typically understood as 'pandemic response' into broader medical, biological and public health questions, as well as to technological and socio-economic implications. Far greater and more diverse sections of national research communities needed to be mobilised. There were fewer instances of 'obvious' candidates to carry out research, and the topics of interest (e.g., air quality, socio-economic effects of lockdown and school closures) required far more agile and wide-reaching approaches to keep pace with increasing and evolving research needs and priorities. Implicitly, these funding activities understand COVID-19 not only as a pandemic affecting public health, but as a much broader and multi-faceted societal emergency.

The emphasis between these two elements of COVID-19 response funding differs among the funders we covered. None of the funders we studied are the sole research funders operating in their respective country, and this has some impact on the shape of their response. For example, in the US, Canada and Germany, other health research funders made substantial investments in response to COVID-19 while the funders we covered focused on other disciplines. The 'pandemic response' element was typically covered by specific health and biomedical research funders, or by equivalent thematic divisions within multidisciplinary funders. The wider 'societal emergency' aspect was typically relevant across the entire disciplinary remit of multidisciplinary national research funders.

As such, all comparator funders supported social science research in addition to biomedical, natural science and engineering research. Some funders only introduced new rapid support measures (e.g., NWO, ZonMw, DFG), whereas other funders (e.g., NRC) were also tasked to deliver new research infrastructure. Some (e.g., UKRI, JST, NRC) introduced or took part in international collaborative programmes, once again in many different thematic and disciplinary domains related to COVID-19 and its wider societal implications.

In this paper, we focus on the wider 'societal emergency' response of multidisciplinary national research funders rather than on the 'pandemic response' of specifically medical research funders or funding divisions.

Figure 1 How funders facilitated rapid funding response processes



Our rationale for this choice is while the latter was by no means an easy or less important task, it was able to draw on prior experience and existing guidance and was limited to a small number of fields and actors suited to carry out the required research activities. The former, by contrast, had little precedent or 'blueprint' and a much broader thematic remit. The lack of precedent, combined with the possibility of future societal emergencies, also means that these 'societal emergency' funding activities beyond 'pandemic response' are particularly likely to include valuable lessons for the future.

In the following sections, we show how funders facilitated rapid funding responses. We distinguish between the processes and activities before and after the point of application submission (Figure 1), as the acceleration mechanisms – and the resulting challenges – are distinct during these two stages: the former pertains to application format and ensuring inclusive participation; the latter pertains to review processes and, most notably, to whether and how peer review can be adapted to the context of a societal emergency and consequent rapid funding deployment.

RAPIDLY RESPONDING TO SOCIETAL EMERGENCY: APPROACHES AND CHALLENGES

DESIGNING AND LAUNCHING THE RAPID RESPONSE PROGRAMMES

All the funders reviewed showed flexibility in their ability to respond to the pandemic quickly. However, we note for context that some funders could draw on prior experience in responding to societal emergencies, meaning they had existing schemes or structures for such purposes. While these may have needed some modification in some cases, other funders needed to create their funding tools from scratch or substantially alter existing schemes that had not been designed for emergency response.

MoST and JST had previous experience responding to natural disasters, pandemics or other crises. ZonMw had experience responding to previous epidemics. Taiwan had already invested significantly in pandemic preparedness after the SARS outbreak. Japan had previously needed to react to major earthquakes and other disasters, so JST used its J-RAPID programme for the emergency response. This, in turn, had been influenced by the NSF's RAPID programme, which had been used in previous emergencies such as Hurricane Katrina. Thus, NSF and JST deployed existing emergency tools, only needing to adjust the programmes for the new emergency at hand.

NRC organised their pandemic response in the framework of the Challenges Programme, creating a specific Pandemic Response Challenge Programme. NRC used the already existing programme mechanisms and did not significantly alter the processes because they believed that using an existing framework would allow them to respond faster than creating a new one.

NRC, MoST and JST also reported fewer problems with implementing the measures, both in terms of organising application preparation and review. This is partly because they used previously tested mechanisms and, as noted above, MoST and JST had overall higher emergency preparedness.

Other funders – NWO and ZonMw, DFG, UKRI – designed new measures or significantly altered existing ones. They reported spending some time on new call design and launch, which often involved some form of formal approval implying lengthier processes, particularly when more than one funder was concerned. For example, NWO and ZonMw started a new joint programme and coordinating between the two funders took slightly more time.

However, funders without ready-made rapid-response instruments reported that management prioritised rapid response over any other activities and accelerated approval processes. Thus, even when funders did not have emergency instruments and had to create new ones, they spent less time approving new measures and starting the operation than would normally be the case for new funding tools.

However, quickly introducing rapid-response measures came at the cost of funders' staff increased workload. Planning and launching calls quickly put significant pressure on the funders' staff. Funders reported significantly increased workload, working overtime and on weekends in remote work circumstances to launch the rapid funding instruments. All consulted funders pointed out this can only happen for a short period and cannot become the norm. Only in countries less affected by the pandemic and with previous experience with responding to a societal emergency (e.g., Taiwan, Japan) did research funders manage to organise the COVID-19 response with less pressure on staff at the design and launch stage.

ACCELERATING APPLICATION PREPARATION AND SUBMISSION

All comparator funders shortened award application timelines, and most of them reduced the length of application forms. The extent to which the time for application preparation was reduced varies among the funders. For example, Dutch health research funder ZonMw reduced the time for application preparation for its interdisciplinary COVID-19 programme from the usual 2-3 months to two weeks. DFG, NRC, MoST and JST allowed longer time periods for application preparation (around one month), but these were still substantially shorter than business-as-usual in all cases.

NWO's Fast-track data programme stands out for its exceptional level of acceleration at this stage: aiming to support data collection for urgent pandemic related research, the programme had a "first-come, first-serve" principle, meaning for example that NWO published the call for applications on a Friday, and by Tuesday the following week, NWO had received enough applications to be able to allocate all budget. NWO programme managers reviewed the applications as soon as they came in. The funder approved applications meeting the minimum requirements of relevance, urgency, expertise, and feasibility.

Two main approaches were evident in terms of reduced application length, which different funders used to varying extents: on one hand, there is the possibility of keeping application form structures the same (i.e. have all the 'usual' sections) but reduce the permitted word or page limits. Most funders followed this approach. On the other hand, there is the possibility of removing some of the business-as-usual application sections altogether. We learned that two funders in our review chose to do this.

DFG and NSF reduced the application length significantly, limiting the usually lengthier applications to just five pages. Other funders opted for a similar but less pronounced approach, reducing the usual length of applications forms only slightly. By contrast, NRC did not make changes of this type, using instead its Challenge Programmes framework to organise its response. Although the programme was new and focused on the pandemic, it used existing Challenge Programme procedures, including application forms. Though not an emergency funding tool as such, NRC deemed the application process and forms for this programme suitable for the urgency of the pandemic. This meant that NRC did not have to create new application forms and reported that this allowed them to launch the call quickly and save time on re-designing or developing new application forms.

Most funders reduced the total length of the applications, not removing specific sections. However, this was not the case for significantly reduced applications. For example, DFG asked for a maximum of five-page applications for its COVID-19 Focus Funding instrument and did not ask applicants to provide information on their track record due. Similarly, in the NSF's RAPID programme form, the key information was largely centred on the proposed research subject.

The rationale for shortening application timelines is self-evident in the circumstances. Likewise, reducing the length and/or detail of applications (both by word/page limits and by removing some of the 'standard' sections altogether) is in part a corollary of this: with shorter time available to applicants, shorter applications ought to help applicants put together an application under such tight constraints. In addition, reviewers have less material to review, which may in theory mean less time spent on reading and assessing applications.

Whilst these steps were almost certainly necessary, our research finds several challenges with shortened applications, both in terms of timelines and application lengths.

One challenge associated with short application deadlines is the quality of the applications and later award implementation. For example, ZonMw observed that some rapidly selected awards later required changes in the project plan because of unanticipated problems during the short application development. DFG reported the quality of applications received for COVID-19 calls was poorer than usual and speculated it might have been due to shorter application preparation time. This was also the case for UKRI's open calls, which attracted a substantial amount of out-of-scope and/or poor-quality applications (alongside many good ones) compared to business-as-usual.

Two funders (DFG and UKRI) experienced challenges with ensuring that peer-review panels had sufficient information to assess applications. As a result, DFG reported that peers sought alternative information resources to find the information that was missing in the applications. Peers needed more time to complete the assessment, and the quality of the additional information peers used could not be assured because it was not provided systematically for all applicants. UKRI staff also commented on the need to regularly source more information from applicants for peer reviewers due to the short application forms used in their Open Call.

The rapid organisation of funding calls also raises concerns about unequal opportunities for some research community members and sustained research career opportunities. NWO reported that fast submission requirements raised concerns in the research community because researchers with care responsibilities could not respond as quickly as others, leaving them at a disadvantage. Elsewhere, it has been reported that female researchers published fewer preprints during the pandemic and started fewer new research projects than males (Viglione, 2020).

In summary, accelerated submission timelines are critical to ensure rapid funding responses and shortening application forms can be a helpful component in this, including for applicants, funder staff and reviewers. However, our research shows that shorter preparation time coupled with lowered barriers to entry may lead to large volumes of applications and lower overall quality of the application pool. This in turn also creates a high burden for the funder's staff and peer-reviewers. Finally, challenges arise to ensure equal inclusion of researchers with difficulties responding quickly.

ACCELERATING APPLICATION REVIEW AND FUNDING DECISIONS

We now turn from the pre-submission to the post-submission phase of COVID-19 response funding. However, we note that several of the features highlighted above also have an effect here on the length or brevity of applications, as well as the large volumes of applications submitted to calls.

Most comparator funders relied on peer-review to assess the applications submitted for COVID-19 response schemes. The main reason for peer-review was (as with regular funding) to ensure scientific quality. At the same time, funders applied new mechanisms and alterations to their usual processes to accelerate peer review. However, several funders also either partly or completely[2] bypassed peer-review in their rapid response.

Funders that least deviated from the traditional peer-review simply instructed peer-reviewers to conduct their reviews in the usual format, but to do so quicker than usual, as JST did in its CREST programme COVID-19 call. This was not the first time JST responded to an emergency in this way, and they already knew they would be able to mobilise the peer community. ZonMw also reduced the peer-review length from 2-3 weeks to receive peer feedback in a few days by simply requesting a fast response from reviewers. Requests for a fast response appear to have been largely effective.

Other funders made more targeted changes in the peer-review process by reorganising and shortening the review process. DFG abolished written panel reviews in the COVID-19 Focus Funding instrument, instead asking peer reviewers to present assessments already written in the panel meeting. DFG also integrated the work of a Review Board with the Grants Commission, usually held separately, saving more time. JST cancelled the joint evaluation meetings between the funders involved in its international J-RAPID programme and instead relied on the assessment provided by partner funders.

We found very few examples of funders making use of two-stage applications (e.g., pre-application followed by a main application). We did not expect this approach to feature strongly in COVID-19 responses due

2 In cases where processes fully by-passed peer review, funders' staff still had an option to involve peer-review to inform their decisions if necessary.

to the extra step taking additional time, despite its common use especially in thematic funding more generally. However, Taiwan's MoST used such an approach, filtering pre-applications to select fewer and better-quality applications that went to full peer review. MoST also organised more panel meetings to speed up decisions.

Finally, three funders by-passed peer-review almost entirely in some of their rapid-response mechanisms. NRC used peer-review for parts of its Pandemic Challenge Programme, but also relied heavily on internal knowledge to assess the applications when needed to speed up the process and support high-risk appetite in its pandemic response programme. This was largely possible because NRC also operates 14 research centres employing scientists and can therefore quickly mobilise relevant scientific expertise.

NSF also relied on its own internal expertise. NSF's RAPID grant mechanism is the only NSF funding mechanism where the funder generally bypasses peer review. It relies on NSF officers for application review and approval. The officers can organise external review if they feel it necessary, but that is not the standard practice. It is designed for quick responses to emergencies, such as when NSF used it in response to hurricane Katrina.

Finally, NWO also made decisions on applications without peer review in its Fast-track data programme. This programme provided small grants and aimed to support rapid data collection during the crisis. NWO therefore decided that its staff should quickly conduct assessments of applications.

The examples of funding instruments that fully or partly bypass peer review are for awards of relatively small sizes. For example, the NWO programme provided maximum grants in the value up to €50k, NRC's programme up to CAN$100k, and NSF RAPID grants were up to US$200k. All programmes asked for short applications. In these examples, funders placed trust in the expertise of their staff (including research centre staff in the case of NRC) and could rely on peer-review as a backup if they encountered difficulties in making the assessment.

As evident from the above examples, research funders used various means to adjust the peer-review process to the urgency of societal emergency. The observed approaches effectively form a 'scale', ranging from making no process modifications and simply speeding up existing processes, via introducing minor administrative efficiencies, simplifying processes (e.g., through the introduction of standing panels), to bypassing peer review almost entirely.

Depending on which approaches they took, funders experienced a range of different and partially interlocking challenges. In varying combinations and trade-offs, these challenges revolved around funder staff and reviewers' workload, the volume of applications received and, critically, the ability to fulfil the requirement of urgency.

Some funders reported that it took too much time to channel funding to awards that had to deliver results very soon. This applies to funding instruments that relied on peer review (or peer review with minor efficiency savings) and saw a high volume of applications – in part as a corollary of shortened application forms and reduced barriers to entry, but also due to the broad thematic remit of calls. For example, DFG cut some steps in the peer-review process and did shorten timelines compared with business-as-usual but was still not satisfied with the length of time the whole process required. UKRI likewise retained peer review (with some efficiency savings in parts of its response) but did not fulfil its ambition of reaching funding decisions within 2-6 weeks on all applications. Such delays were not entirely a result of peer review itself, but also of the need to quickly process a large volume of applications in this way.

Such time delays are not evident in instruments that simply asked reviewers to return their feedback much faster. The three funders who relied on this approach (JST, MoST, ZonMw) did not report any significant problems or failure to meet the objectives related to urgency. However, even at smaller scale, this approach was deemed unlikely to be sustainable as it implies a heavy workload for peers and was noted not to be feasible for very large funding instruments with many applications.

Funders that bypassed peer-review did not report any problems with not meeting the urgency objectives and believed[3] they made funding decisions faster without, rather than with, full peer-review. However, as noted, bypassing peer review was only practiced for small-sized awards.

Related to the above is the issue of funders' staff workload associated with managing large numbers of applications and accelerated peer-review processes involving recruiting peers and repeated requests to peer-reviewers to return their assessments. The same also applies to funders who rely on their staff for application assessment mainly because rapid response mechanisms received many applications. To address this, NSF requested the applicants to contact NSF officers before submission to ascertain if their application would be appropriate. Still, even with this procedure, NSF received thousands of applications.

In short, funders faced the most serious challenges when trying to accommodate large application volumes (brought about by broad topic remit and shortened applications), conduct more-or-less full peer review on most or all applications, and ensure rapid funding deployment (as demanded by the situation). Generally, it was the latter issue on which funders fell short in such cases. However, substantial workload and stress levels for reviewers and/or funder staff also occurred in instruments where at least one of these three parameters (volume, speed, peer review) was removed.

URGENCY VERSUS QUALITY? UNPACKING THE DICHOTOMY

Research funders accelerated funding mechanisms throughout the whole funding process. Table 2 presents an overview of the accelerated funding mechanisms we identified - from shortened pre-application timelines to expedited peer-review and summarises the associated advantages and potential hazards.

3 Note that evidence on the NSF RAPID instrument is based only on desk review.

Table 2 Funding mechanisms at a glance – advantages and disadvantages

Mechanism	Main Advantages	Main Hazards
Shorten timeline from call launch to submission	Shortening of the overall funding process	May result in poorer quality proposals May exclude individuals with caring responsibilities or otherwise unable to respond
Shorten application form (lower permitted lengths and/ or remove sections)	Eases applicants' ability to write applications in short times available May enable faster review	Lowers 'barriers to entry', potentially leading to large volumes of applications May lead to information gaps for reviewers
Expression of interest or pre-application phase prior to full application	Lowers the volume of applications going to peer review Increases relevance of the pool of applications	Takes additional time May mean substantial workload for funder staff or standing panels in charge of 'sifting'
Full peer review of standard-length applications	Optimally safeguards scientific quality and standards	May either take a long time or require substantial pressure on reviewers Not suitable for urgent funding in conditions of high application-influx
Simplified decision-making process (e.g., combining/by-passing some decision-making bodies)	Leads to minor time/efficiency savings May slightly reduce administrative burden	May not be suitable for large award sizes or funding decisions that require strategic oversight
Modified peer review (e.g., standing panels only, no individual remote peer reviews)	Leads to some time savings May reduce funder staff burden to identify remote reviewers	Substantial pressure on standing panels, especially in cases of high application influx
No peer review (or in exceptional circumstances only) – decision by funder staff	Substantial time savings No or minimal administrative burden to identify/ organise external peer or panel reviews	Potential lack of process-trust from the research community (or requires trust in funder staff) Generally only deemed feasible for small awards

Pre-submission process alterations allowed to save time but also caused some challenges. Key challenges were around proposal quality, information gaps for reviewers and large volumes of applications. Post-submission peer-review is where we saw most change, challenges and opportunities. Thus, we discuss this in more detail onwards.

Traditionally, peer review has been the default mechanism to make decisions in research grant funding. Specifically, a sequence of external peer reviews followed by ranking and sorting of applications by a review panel is in use at almost all research funders across the globe, be it for basic research funding, thematic funding, or innovation-oriented funding. The research community places a great deal of trust in peer review, and while the 'peer review burden' has been acknowledged for some time (Guthrie et al, 2013; Herbert et al, 2015; Schroter et al, 2010), the use of peer review to allocate funding does not present an operational difficulty.

This changed in the context of the 'societal emergency' funding activities conducted in response to COVID-19. The range of process decisions taken by the funders signal a perceived tension between the need for urgency on one hand and the need to conduct the fullest possible peer review on the other. If we understand peer review as a central mechanism for scientific quality assurance, we can simplify the central tension to 'speed versus quality'. This tension is exacerbated when dealing with large volumes of applications.

Several interviewees for our study acknowledged this perceived tension and indeed, the funders we reviewed responded to this tension in several different ways – in some cases, the same funder managed it differently in different funding instruments. Regardless of the approach taken, managing the tension between urgency and the need for peer review typically resulted in at least some personal cost in the form of stress and high workloads, either to funder staff or to reviewers, or to both.

This tension highlighted in funders' COVID-19 responses illuminate some long-lasting issues with research funding. One is the debate about the quality of peer-review. Scholars have pointed to the lengthy processes it involves (Guthrie et al, 2013), and there is growing evidence that peer reviewing all applications or relying exclusively on peer review does not necessarily lead to optimal funding outcomes. Peer review may lead at least in part to arbitrary outcomes especially in conditions of high application volumes and low success rates (Abdoul et al, 2012; Clarke et al, 2016; Graves et al, 2011; Mutz et al 2016). The urgency of 'societal emergency' funding thus provides grounds to question whether peer review should always feature.

The literature also shows that peer review may be biased against risk, i.e., putting especially innovative and 'transformative' ideas at a disadvantage (Guthrie et al, 2018; Langfeldt, 2006; Nuffield Council on Bioethics, 2014). For example, Franzoni et al (2021) discuss peer-review aversion and, in light of the pandemic, illustrate how Katalin Karikó, a scientist who conducted pioneering research related to mRNA-based drugs, did not succeed with her early applications for funding because their research was considered too preliminary and risky.

Problems with lengthy processes and risk aversion became evident also in some aspects of the pandemic response. It raises the question of whether traditional full peer review is always compatible with rapid response. Complete or partial bypassing of peer-review might also be rel-

evant to open the doors for risky and potentially breakthrough research if that is the desire of the funder. 'Pandemic responses' in the conventional sense (see above) may not be the right place to contemplate high-risk funding. But 'societal emergencies' more generally might in part require riskier solutions to complex and novel problems.

Looking across the experiences of funders covered by our study, the speed and quality assurance are not a straightforward dichotomy, or even part of a one-dimensional 'scale'. The range of processes and modifications used across the seven funders highlight that there is a range of levers that may be combined in many ways. These offer a starting point to constructing generic emergency response toolkits that may be drawn upon in the future.

CONCLUSION: TOWARDS A RAPID-RESPONSE TOOLKIT FOR NATIONAL RESEARCH FUNDERS

Research funders are typically understood to have up to three 'missions': first, to fund basic, curiosity-driven research, bottom-up (researcher driven) and in the shape of projects and fellowships. Second, to fund innovation related activities (this especially applies to combined R&I funding agencies). More recently, research funders have also taken on thematic missions, aiming to fund research relevant to solving societal challenges, for example in relation to the UN Sustainable Development Goals and through 'research for development' programmes. We note that these missions are sedimentary: the presence of new missions does not make older ones less important but rather expands the range of funders' activities. The importance of funding basic research without top-down thematic imperatives is well established (see e.g., Kohse-Höinghaus et al., 2019) and we concur with such sentiments.

'Societal emergency' funding may present a fourth 'mission' for funders: though essentially oriented to solving a societal challenge, it is distinct from thematic funding due the extreme urgency characterising its deployment.

Large amounts of funding reached researchers faster than usual owing to process modifications. Some have raised questions whether the introduced changes can be transferred into the everyday operation of funders (OECD, 2021b; Wilsdon, 2021). Given the challenges experienced by funders, we likewise find it unlikely that many of the approaches taken for COVID-19 response funding are appropriate for 'business-as-usual'. For this reason, we deem it most appropriate to understand 'societal emergency' funding as a distinct activity requiring distinct processes.

Funders across the globe mobilised to respond to the societal crisis brought about by COVID-19. For many of them, this was the first time they needed to deploy funding rapidly in this way. Our assessment highlights that rapid funding requires adaptations to the usual funding process to ensure that research can produce impact within helpful timeframes.

In cases of societal emergency, our findings highlight a (non-exhaustive) selection of levers and techniques available to funders. Some of them rule each other out, others do not. Not all are useful for all types of awards, so if we contemplate a toolkit or guidance for 'societal emergency' funding, we need to consider:

- The size (monetary value) of awards
- The level of urgency (these might differ depending on the nature of the crisis or which aspect of the crisis is sought to be addressed by the funding)
- The thematic breadth of the call (likely a determinant of the volume of applications)
- The level of 'risk-appetite' (e.g., based on the need for especially innovative solutions in what may be uncharted territory)

Societal emergencies may require several different types of awards, suggesting that funders need to have a range of funding tools at their disposal and systematically use the 'levers' of topic urgency, risk appetite, award size and, ultimately, internal knowledge and expertise to make rapid decisions where feasible.

The experiences of NSF and JST illustrate the value in having a purpose-made rapid response funding instrument ready for use for societal emergencies. The NSF RAPID and JST J-RAPID programmes allowed both funders to use the institutional knowledge and previously tested processes to mobilise for the COVID-19 pandemic. However, the tensions and choices described in this paper suggest that a ready-made suite of funding tools for societal emergencies might be even more useful. While we do not wish to be prescriptive (full lessons from COVID-19 funding responses across the globe have yet to be drawn), our findings allow us to posit as a generic model of three scheme types – to be deployed with varying emphases depending on the nature of a societal emergency and its consequent research-needs:

1. **An instrument to fund awards as rapidly as possible, using minimal or no peer review:** this may be reserved only for the most urgent research-needs that need to be deployed within days rather than weeks (e.g., rapid data collection needs to monitor a particular aspect of an unfolding crisis). Thematic remit ought to be relative tightly defined and informal enquiries or expressions of interest possible, to limit the influx of out-of-scope applications. Internal funder staff to review and take funding decisions, with additional experts to be consulted informally if required. Other than in exceptional circumstances, awards on this scheme would be of relatively low financial value. This instrument may ideally have one or more topic-specific and highly time-bound calls rather than being a rolling open call throughout the crisis. Because of urgency, short applications, and a shorter than usual timeline for submission, would be reasonable

2. **An instrument to fund awards rapidly using simplified or modified peer review:** societal emergencies might have a broad and multi-disciplinary range of research-needs that are urgent but can countenance a few weeks of waiting-time. This scheme should be designed to accommodate a high intake of applications but process them relatively quickly. This could be facilitated through a 'sifting' stage where 1–2-page pre-applications (or summary sections of full applications) are rapidly sifted for relevance by funder staff or by a standing panel, so that the volume of applications going to full peer review is limited. Full applications and application timelines may be shorter than they would be for equivalent-sized business-as-usual awards. Funders may consider using standing panels to conduct reviews, relying on

additional expert reviews only in cases where panels do not have sufficient thematic knowledge to make judgements. This instrument may have one or more topic-specific and time-bound calls, or may take the shape of a rolling open call with a loosely delineated (and potentially evolving) topical remit

3. **An instrument to fund large, strategic awards relatively rapidly using full peer review:** where research needs are pressing but not immediate, larger awards (e.g., for centres, facilities, and major consortia) may require full external peer and panel review. For such instances, this instrument will most closely resemble a funder's business-as-usual processes, including full-length applications, both to ensure high standards of scientific quality assurance and to heighten barriers to entry (i.e., reduce application influx). Pre-applications or applications by invitation only may be considered for this instrument. Peer reviewers may be briefed that reviews in this scheme constitute exceptional circumstances, meaning that peer reviews need to be returned within a much shorter time than usual. Public agencies other than the research funder may also provide input into decisions on strategic investments (e.g., as was done by the Scientific Advisory Group for Emergencies [SAGE] for UKRI's response)

As with any other funding measure, rapid-response mechanisms should remove barriers preventing all researchers from contributing. It can be challenging to balance the urgency and need for quick submission of applications. Still, funders can introduce flexible policies to accommodate the needs of the research community, train staff or peers involved in assessing the applications to assess applicants equitably or provide support to cover care costs.

We stress that the above is an initial suggestion and primarily intended as an illustration of the combination of best practices revealed in our research. Experiences of the many funders not covered by our study may yield additional insights leading to substantially different models. Further, as impact evaluations of various funders' COVID-19 responses take place and reach the public domain, additional insight will be gained into what kinds of funding instruments and process modifications produced the most relevant, consistent and/or innovative results.

'Societal emergency' funding may become a new occasional mission for research funders. The funders who participated in our study must be lauded for their efforts and thanked for their participation. Whether COVID-19 was their first societal emergency response or not, our research found ample markers of good practice, and we offer our findings as a first step towards easing the burden of any future crises that may come.

REFERENCES

Abdoul, H., Perrey, C., Amiel, P., Tubach, F., Gottot, S., Durand-Zaleski, I., & Alberti, C. (2012). Peer review of grant applications: criteria used and qualitative study of reviewer practices. PLoS One, 7(9), e46054, DOI: 10.1371/journal.pone.0046054

Franzoni, C., Stephan, P., Veugelers, R. (2021). Funding Risky Research. NBER Working Paper No. 28905. Available: https://www.nber.org/system/files/working_papers/w28905/w28905.pdf

Graves, N., Barnett, A. G., & Clarke, P. (2011). Funding grant proposals for scientific research: retrospective analysis of scores by members of grant review panel. Bmj, 343, d4797, DOI: 10.1136/bmj.d4797

Guthrie S., Ghiga I. & Wooding S. (2018). What do we know about grant peer review in the health sciences? [version 2; peer review: 2 approved] F1000Research, 6:1335, DOI: 10.12688/f1000research.11917.2

Guthrie, S., Guerin, B., Wu, H., Ismail, S., & Wooding, S. (2013). Alternatives to peer review in research project funding. Prepared for the UK Department of Health-funded Centre for Policy Research in Science and Medicine (PRiSM). Rand Europe. Available: https://www.rand.org/content/dam/rand/pubs/research_reports/RR100/RR139/RAND_RR139.pdf

Herbert, D. L., Graves, N., Clarke, P., & Barnett, A. G. (2015). Using simplified peer review processes to fund research: a prospective study. BMJ open, 5(7), e008380, DOI: 10.1136/bmjopen-2015-008380

JST (2021). Japan Science and Technology Agency. Available: https://www.jst.go.jp/EN/about/pdf/outline_e.pdf

Kohse-Höinghaus, K., Fuchs H., Zhang T., Wu Y., (2019). Explore the Unknown—The Value of Basic Research. Angew. Chem. Int. Ed. 58(50), 17882, DOI: 10.1002/anie.201913851

Kolarz, P., Arnold, E., Bryan, B., D'hont, J., Horvath, A., Simmonds, P. & Vingre, A. (2021). Process review of UKRI's research and innovation response to COVID-19. Available: https://www.ukri.org/wp-content/uploads/2022/01/UKRI-180122-ProcessReviewUKRIResponseCOVID19-FinalReport.pdf

Langfeldt, L. (2006). The policy challenges of peer review: managing bias, conflict of interests and interdisciplinary assessments. Research evaluation, 15(1), 31-41, DOI: https://doi.org/10.3152/147154406781776039

Lurie, N., Saville, M., Hatchett, R., Halton, J. (2020). Developing Covid-19 Vaccines at Pandemic Speed. New England Journal of Medicine, Vol. 382/21., DOI: http://dx.doi.org/10.1056/nejmp2005630

Mills, A. J., Durepos, G., & Wiebe, E. (2010). Encyclopedia of case study research (Vols. 1-0). Thousand Oaks, CA: SAGE Publications, Inc. DOI: 10.4135/9781412957397

MoST (2021). About MoST. Available: https://www.most.gov.tw/folksonomy/list/472e87f2-a8cc-41f5-94f8-c68c147840c4?l=en

Mutz, R., Bornmann, L., & Daniel, H. D. (2016). Funding decision-making systems: An empirical comparison of continuous and dichotomous approaches based on psychometric theory. Research Evaluation, 25(4), 416-426, DOI: 10.1093/reseval/rvw002

National Research Council Canada (2021). About the NRC. Available: https://nrc.canada.ca/en/corporate/about-nrc

National Science Foundation (2021). About the National Science Foundation. Available: https://www.nsf.gov/about/#:~:text=The%20Na-

tional%20Science%20Foundation%20(NSF,people%20to%20create%20knowledge%20that

Nuffield Council on Bioethics. (2014). The findings of a series of engagement activities exploring the culture of scientific research in the UK. Nuffield Council on Bioethics. Available: https://www.nuffieldbioethics.org/assets/pdfs/The-culture-of-scientific-research-report.pdf

NWO (2021). About NWO. Available: https://www.nwo.nl/en/about-nwo

OECD (2021a). STIP COVID-19 Tracker. Available: https://stip.oecd.org/covid/

OECD (2021b). OECD Science, Technology and Innovation Outlook 2021. Times of crisis and opportunity. Available: https://www.oecd-ilibrary.org/science-and-technology/oecd-science-technology-and-innovation-outlook-2021_75f79015-en

Oliveira, J. F., Pescarini, J. M., Rodrigues, M. S., Almeida, B. A., Henriques, C., Gouveia, F. C., Rabello, E. T., Matta, G. C., Barreto, M. L., & Sampaio, R. B. (2020). The global scientific research response to the public health emergency of Zika virus infection. PloS one, 15(3), e0229790. https://doi.org/10.1371/journal.pone.0229790 Viglione, G. (2020). Are women publishing less during the pandemic? Here's what the data say. Nature 581, DOI: https://doi.org/10.1038/d41586-020-01294-9

Schroter, S., Groves, T., & Højgaard, L. (2010). Surveys of current status in biomedical science grant review: funding organisations' and grant reviewers' perspectives. BMC medicine, 8(1), 62, DOI: 10.1186/1741-7015-8-62

Stoye, E. (2020). How research funders are tackling coronavirus disruption. Available: https://www.nature.com/articles/d41586-020-01120-2

Wilsdon, J. (2021). The post-pandemic politics of research prioritisation. Keynote presentation at STI 2021 conference. Available: https://www.youtube.com/watch?v=YV6mq8o5Hrc

AUTHORS

ANETE VINGRE
Technopolis Group
2-3 Pavilion Buildings, Brighton, UK, BN11EE
E: anete.vingre@technopolis-group.com

PETER KOLARZ
Technopolis Group
2-3 Pavilion Buildings, Brighton, UK, BN11EE
E: peter.kolarz@technopolis-group.com

BILLY BRYAN
Technopolis Group
2-3 Pavilion Buildings, Brighton, UK, BN11EE
E: billy.bryan@technopolis-group.com
ORCID ID 0000-0002-5771-2354

KEYWORDS: research funding, rapid research, R&I policy, COVID-19, peer review

EFFECTS OF COVID19 PANDEMIC ON R&D FUNDING SCHEMES IN GERMANY – FIRST RESULTS OF A COMPARATIVE ANALYSIS OF EMPIRICAL DATA

CHRISTIANE KERLEN, SONJA KIND, KAROLINE RODRIGUEZ, LEO WANGLER, GUIDO ZINKE AND JAN WESSELS
DOI: 10.22163/fteval.2022.539

ABSTRACT

In 2020 and 2021, practically all areas of work and life were under the influence of the Corona pandemic. In the course of 2020, it was already apparent that the consequences of the Corona crisis would also have a considerable impact on grant recipients and their projects and thus on the implementation and goal achievement of the funding programmes, and that this would be reflected in the results of evaluations. From the perspective of an evaluation of innovation funding programmes, the Corona pandemic represents an external influencing factor that could negatively affect the intended impact of a funding programme with regard to the successful implementation of innovation projects and behavioural change of actors towards innovation orientation. This article provides initial answers to the questions of what effects the Corona pandemic has on innovation promotion, what adaptation strategies can be observed and how these can be provisionally evaluated. For this purpose, the authors of this paper draw on current, as yet unpublished data from evaluation surveys in 2020 and 2021.

1 INTRODUCTION

In 2020 and 2021, practically all areas of work and life were under the influence of the Corona pandemic. The effects on society, the economy, teaching, research and administration were and are enormous. With regard to innovations, there are signs of an ambivalent development. On the one hand, numerous innovations were triggered and there was a significant surge in digitalisation, but on the other hand, innovation processes were also slowed down, and many central innovation processes promoted within the framework of innovation funding programmes were inhibited.

In the course of 2020, it already became apparent that the consequences of the Corona crisis would also have a considerable impact on funding recipients and their projects, and thus on the implementation and goal achievement of the funding programmes, and that this would be reflected in the results of evaluations (Kerlen 2020).

For many actors in research projects, the situation was very challenging because, for example, access to laboratories was temporarily restricted or no longer possible due to a lack of staff, interrupted supply chains, meetings cancelled, conferences not held, funding diverted to corona measures, lack of technical equipment for communication, etc. Due to the consequences for individual sectors and the mostly difficult financial situation of the municipalities, it became apparent that technology policy goals of innovation promotion could no longer be achieved in all cases.

For the evaluation of research, technology and innovation policy measures, the question arises as to how the "Corona crisis" can be adequately captured. From the perspective of an evaluation of innovation funding programmes, the Corona pandemic represents an external influencing factor that could negatively affect the intended impact of a funding programme with regard to the successful implementation of innovation projects and behavioural change of actors towards innovation orientation.

This is relevant insofar as the impact of a programme affected by the Corona pandemic measured by an evaluation could thus be lower in comparison to other programmes that were not implemented during the Corona pandemic, without this being attributable to the programme itself. In addition, the pandemic has led to adjustments and changes (e.g., other forms of communication, extensions of the programme duration, new thematic focuses) in the projects, which in turn are likely to have both intended and unintended effects.

This article provides initial answers to the questions of what effects the Corona pandemic has had on innovation funding, what adaptation strategies can be observed and how these can be provisionally evaluated. For this purpose, the authors of this paper draw on current, previously unpublished data from evaluation surveys in 2020 and 2021. Five evaluations and monitoring processes of innovation funding programmes of the Federal Ministry of Education and Research (BMBF), the Federal Ministry for Economic Affairs and Climate Action (BMWK) and the Federal Ministry of Transport and Digital Infrastructure (BMVI), which have been running since summer 2020, are examined. In the surveys, current and former grant recipients as well as experts were asked about the impact of the Corona pandemic. In total, data from eight survey rounds are available. The main motivation for the selection of the measures under consideration was the availability of data within the framework of ongoing evaluations.

The article presents the first results of these surveys as examples. Furthermore, a possible impact model is proposed that takes into account the external shock "Corona" on ongoing innovation processes. Finally, possible further research needs and strategies are presented.

2 INFLUENCE OF PREVIOUS CRISES ON THE INNOVATION SYSTEM AND CURRENT PANDEMIC-RELATED TRENDS

The discussion on the possible impact of the pandemic on research and development (R&D) is based not only on current observations but also on findings from previous crises. The experience of the 2008/2009 financial crisis in conjunction with observations of reactions to it suggest that private R&D expenditure tends to be reduced procyclically in times of crisis. The reasons for this are primarily short-term financing bottlenecks of companies as well as uncertain expectations regarding market developments (Dachs / Peters 2020). As a result, fewer contracts are awarded to research service providers such as Fraunhofer institutes, universities and universities of applied sciences. Furthermore, there is an overall withdrawal from R&D projects (Azagra-Caroa et al. 2020). For German research institutions, this means a noticeable decline in third-party funding, at least temporarily, with corresponding consequences for funding opportunities and staff continuity (Estermann 2020).

However, the experiences of the financial crisis also speak for an anti-cyclical innovation behaviour of a few companies and thus for a further differentiation of the corporate landscape (Dachs / Peters 2020). In addition, studies show that large international technology groups in particular tend to come through the crisis better than smaller companies (Economist 2020). Overall, these trends are likely to lead to a further spread of the national and international corporate landscape into a few successful, highly innovative technology champions and the broad mass of less innovative and thus less productive companies in the long term. This poses a risk to the competitiveness of the German economy in the medium to long term.

The effects are likely to be particularly serious for small and medium-sized enterprises (SMEs). Scientific study results suggest that a withdrawal from research is difficult to reverse, especially for SMEs, as it goes hand in hand with the dismantling of relevant internal structures and competences and the access barriers for re-entry become higher (Rammer / Schubert 2016). The declining innovation orientation in the breadth of German SMEs, which was already observed before the Corona pandemic, could be further intensified by this.

By mid-2021, numerous data on the effects of the pandemic on the innovation system were already available. For example, various indicators (number of patents, innovation expenditure) and industry data point to a slight decline in innovation activities (European Patent Office 2021, ZEW 2/2021). According to a survey by the German Centre for Economic Research (ZEW), companies in Germany expect a slight decline in innovation expenditure of 2.2 per cent in 2020 compared to 2019. Although the decline is significantly smaller than in the financial crisis (decline of eleven percent in the financial crisis of 2009), it is more pronounced for smaller firms than for large ones (ZEW 7/2021).

Start-up activities were also affected by the pandemic and generally declined more in 2020 than in previous years (KfW 2021). In contrast, the financing of innovative start-ups in Europe has tended to improve despite the crisis (EY 2021). The number of financing rounds in Germany rose by 62 percent to 588 in the first half of 2021, reaching a new record level. The total value of financing rounds also jumped to €7.58 billion (EY 2021). Overall, however, it is apparent that young innovative companies have been hit harder by the crisis in comparison (ZEW 11/2020).

Meanwhile, data is also available on the consequences for research and innovation funding. For example, the German Research Foundation (DFG) recorded an overall decline in research proposals in basic research for 2020 (DFG 2021). The same applies to application-oriented research funding programmes, which address not only research institutions and universities but also companies.

In contrast, figures for the start-up support programme EXIST show that interest in funding in the area of start-up support continued unabated even during the Corona pandemic (Munich Startup 2021). In 2020, 426 applications were received for the EXIST start-up grant. This was a peak since the programme was launched in 2007. This could indicate a stronger focus on self-employment entrepreneurship during the crisis.

3 IMPACT MODEL: INTEGRATION OF THE COVID19 EFFECT AS AN EXTERNAL INFLUENCING FACTOR

Most evaluations are based on a programme theory. The starting point of an evaluation is therefore usually an impact model of the measure under investigation and its environment. As a rule, external as well as internal factors of influence are taken into account, including the Corona pandemic. In order to be able to take this new effect into account appropriately, an impact model was developed that can be applied in a generalised form to different funding measures in the policy field of research, technology and innovation policy. It traces typical impact paths and highlights the influence of Covid19 (Figure 1).

It can be assumed that economic slumps caused by the pandemic lead to a worsening of the economic situation of subsidised companies, which in turn change their innovation behaviour and focus more on existing business, which stabilises turnover in the short term, instead of investing in future turnover within the framework of R&D projects. Furthermore, there is an assumption that resource bottlenecks caused by border closures in the first peak phase of the pandemic in spring 2020 lead to delays in project implementation. The delays in turn have a negative impact on project utilisation, e.g., because the product launch on the market is delayed.

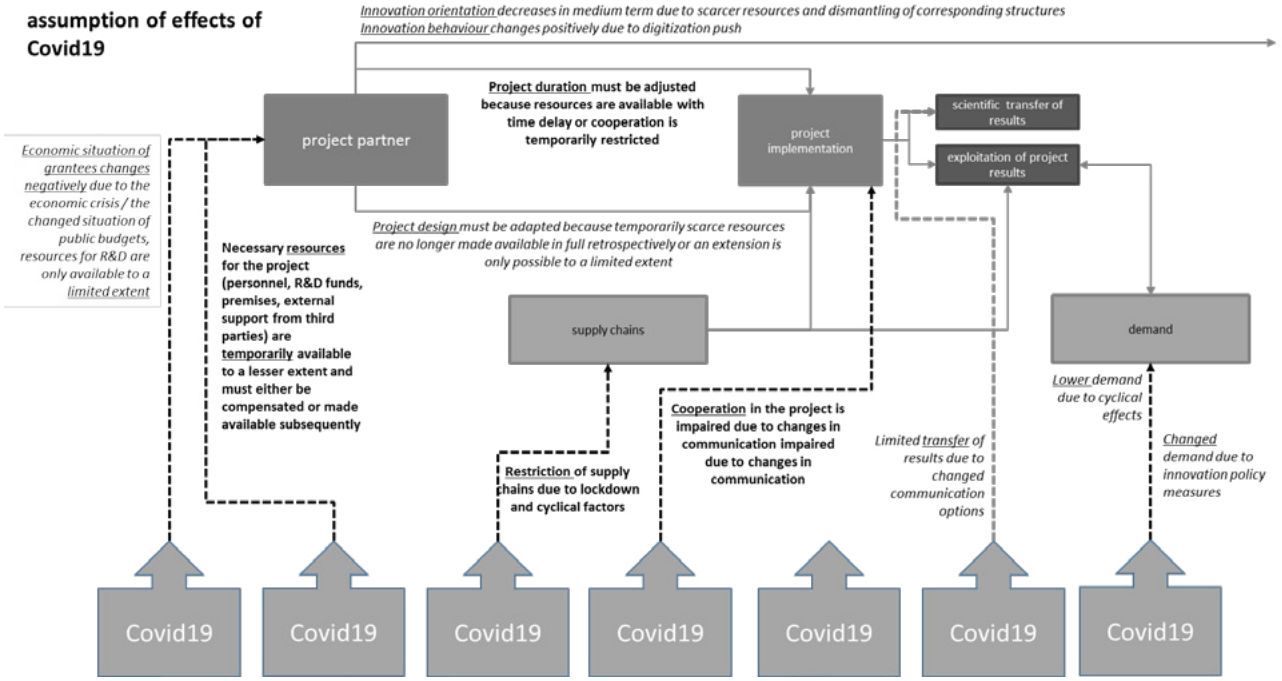
The following primary cause-effect relationships can be assumed:

- Due to temporarily reduced internal resources, interrupted supply chains and/or limited communication in the project team, some projects take longer or require more resources.
- If these temporary constraints cannot be compensated for in the course of the project, the project design of some projects will be changed, potentially leading to less successful or less "effective" projects.
- The markets and thus demand also change due to cyclical effects and/or additional Corona-induced political measures (e.g., in favour of future-oriented technologies such as electromobility), which in turn influences the impact of the projects and thus of the overall programme.
- In addition, the innovation behaviour of the actors could change during the crisis. In the medium term, changes in the availability

of resources could have a negative effect, whereas the digitalisation boost experienced during the crisis could have a positive effect.

The effects that have already been proven to some extent by initial data are shown in bold, while the effects that have not yet been empirically proven are assumed in italics.

Figure 1: Impact model - influence of the Corona pandemic on R&D projects



4 EMPIRICAL FINDINGS OF COVID19 PANDEMIC EFFECTS ON GERMAN R&D FUNDING SCHEMES

The effects of the Corona pandemic on grantees in several innovation funding programs are shown below. What all programs have in common is their focus on applied research projects that are usually carried out by a consortium of project partners from both academia and the industry. Most funded projects deal with technological innovations. The programs are presented in anonymized form (A to E), and are the responsibility of the following German federal ministries:

- Federal Ministry of Education and Research (BMBF).
- Federal Ministry for Economic Affairs and Climate Action (BMWK)
- Federal Ministry of Transport and Digital Infrastructure (BMVI)

The authors of the study carry out evaluation and monitoring activities of all programs on behalf of the respective ministries. To assess program implementation, achievements and impact, funding recipients have been addressed with online surveys. In order to measure the corona effects, a set of specific questions on possible effects due to the corona pandemic was integrated into the surveys conducted.

The surveys took place at two points in time in summer 2020 and 2021. For one programme, there was an additional survey wave in spring 2021.

Programme E differs from the other programmes in that experts were interviewed rather than funding recipients. The questions were therefore not directed at the effects of the Corona pandemic on an individual funding project. Instead, the influence on innovation in general was considered.

The results presented must be interpreted against the background of the course of the pandemic. The intensity of the pandemic had eased somewhat in the summer of 2021 compared to the previous months, but the respondents from the survey wave in the summer of 2021 answered against the background of a severely restrictive lockdown in the winter of 2020/21. In addition, the expectation was that the coming winter could again lead to an increase in incidence levels and renewed restrictions on public life. In contrast to the first survey in the summer of 2020, one year later the respondents had already gained experience in dealing with the pandemic over a period of months; in addition, various adaptation strategies, especially through the digitalisation of processes and communication formats, had already been implemented for some time.

It should be noted that the questions asked were not the same for all programmes but were adapted to the specific features. This means that a direct comparison between the programmes is only possible to a limited extent.

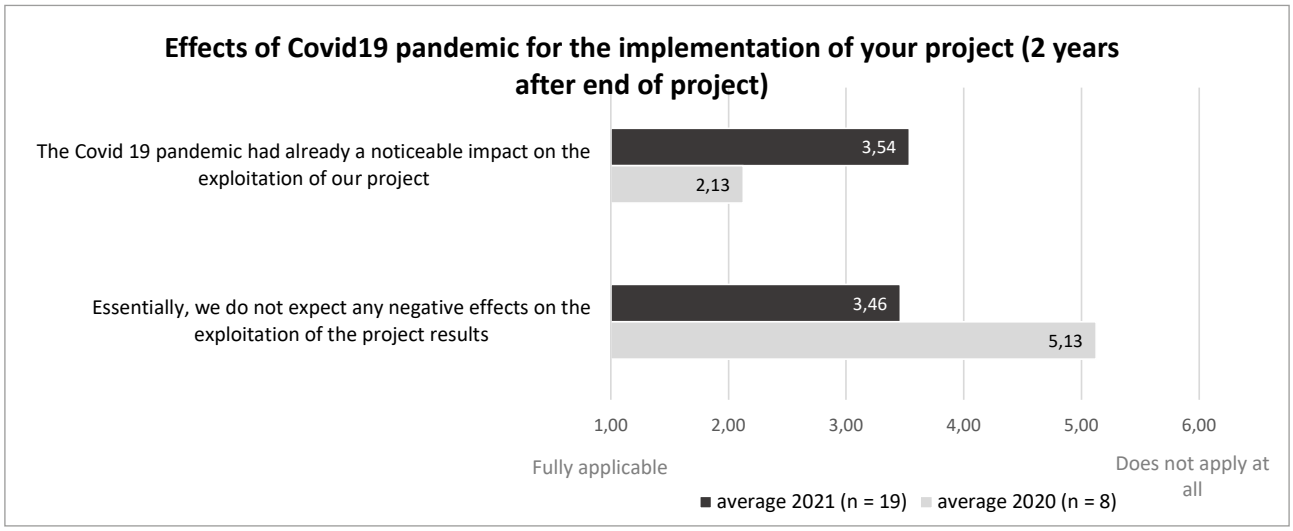
4.1 PROGRAMME A

For Programme A, a comparison of the two surveys in 2020 and 2021 shows that the assessments of the severity of the impact of the pandemic on the projects in the following year 2021 are mostly more pronounced than a year earlier.

For projects still in progress, both negative consequences (e.g., limited staff availability, postponement of work packages) and positive effects (e.g., increasing demand, added value through the use of versatile digital formats and communication tools) are mentioned more frequently in 2021 than in 2020. This observation can be interpreted to mean that the responses to the 2020 survey were still borne by a certain optimism from the early days of the pandemic, and there was hope that the effects would not be so severe. One year later, there were already concrete experiences from the second half of 2020 and the first half of 2021, so that both positive and negative effects were weighted more heavily. In addition, it can be concluded from the data that although adaptation strategies of digitalisation were considered successful and positive to a certain extent, fundamental challenges such as limited personnel availability could not be sufficiently compensated.

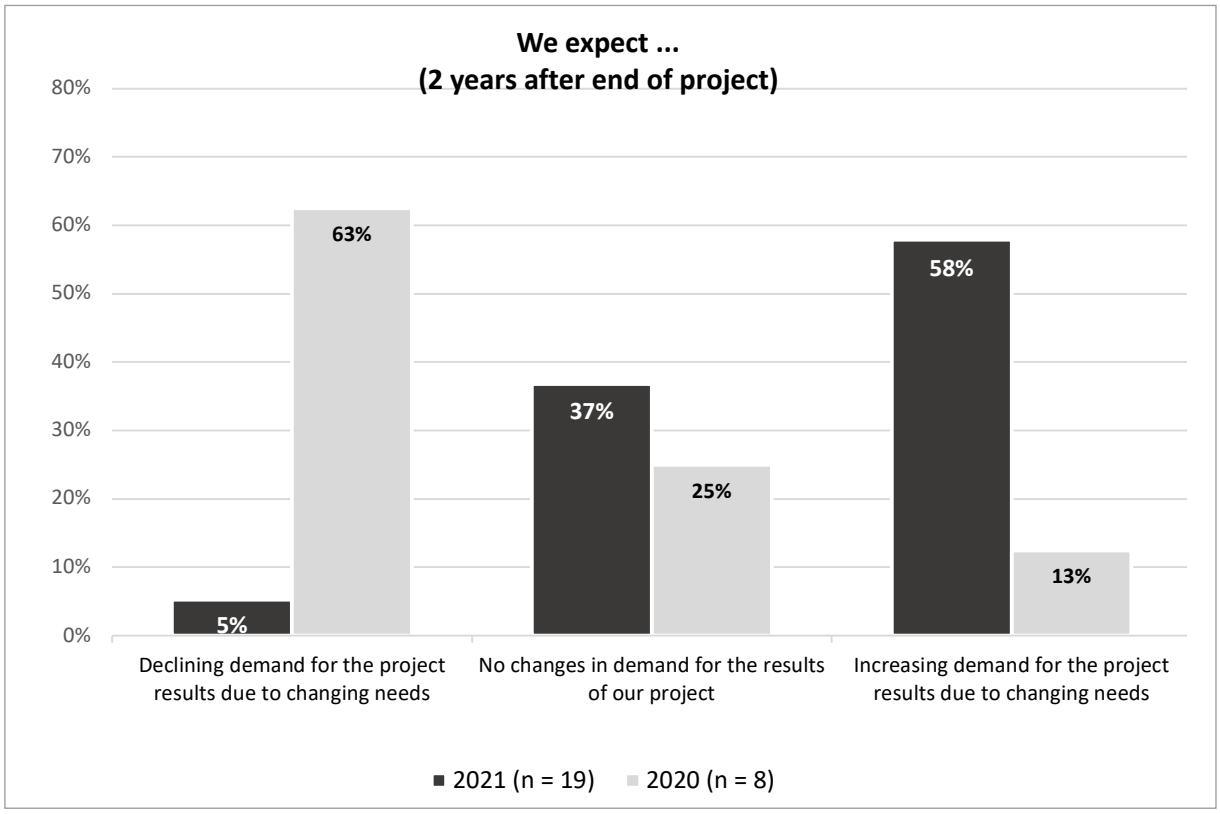
Grantees whose projects had already been completed for some time and were in the exploitation phase assessed the impact of the Corona pandemic in part significantly differently from those respondents who had to implement their projects during the pandemic (figure 2). Anticipated negative impacts of the pandemic were rated as lower in 2021 compared to 2020 by respondents who completed their project about two years ago. Respondents from projects completed five years ago anticipated negative impacts somewhat differently. This group estimated that the potential negative effects would be somewhat lower in 2021. This could be due to the fact that the group in the 2020 survey with project completion more than five years ago had already completed all major implementation steps, while the projects in the 2021 survey wave had only been completed for about 1 to 1.5 years at the beginning of the pandemic and were still more in the process of implementation. However, less negative effects were expected in 2021 than in 2020.

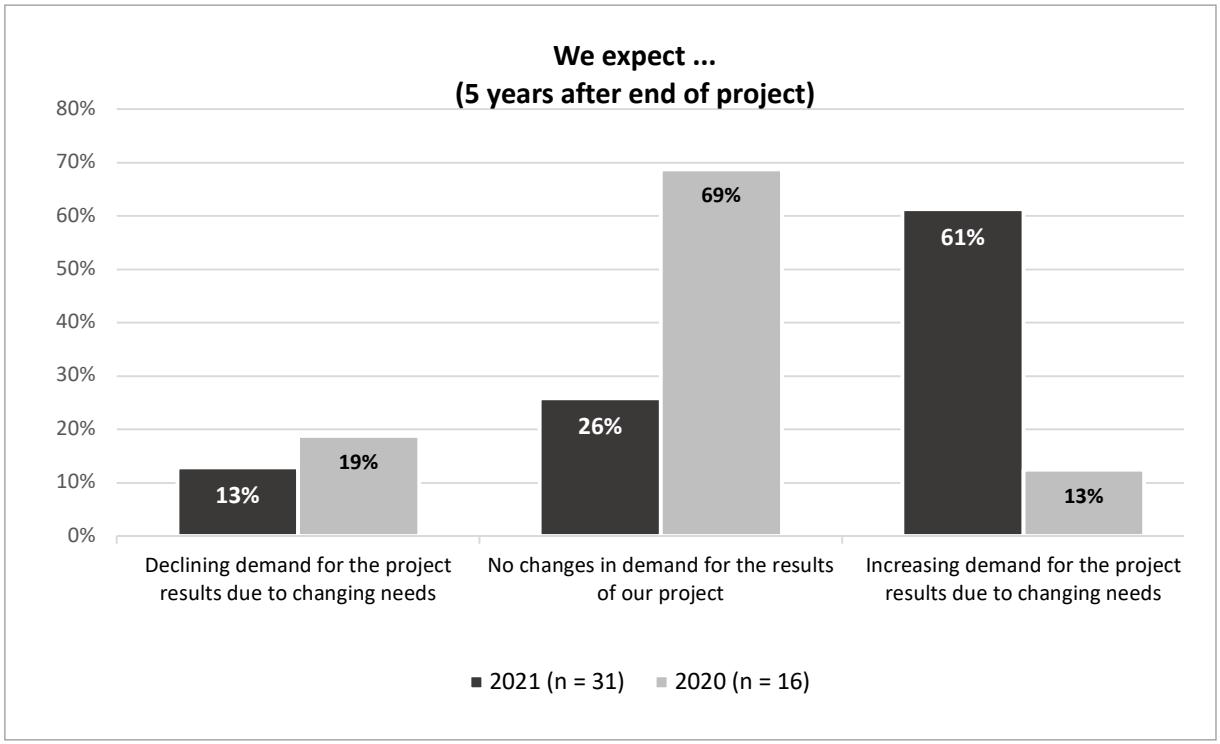
Figure 2: Impact of the CORONA pandemic on project implementation (2 vs. 5 years after project completion)



In 2021, the respondents are much more optimistic about the potential demand for their project results compared to the previous year. Whereas in 2020 the projects in the 2-year follow-up survey still predominantly expected falling demand, this assessment turns towards rising demand in 2021. A similar change can be observed in the projects with a greater distance to the end of the project (5-year follow-up survey).

Figure 3: Expectations of demand for project results (2 vs. 5 years after project completion)

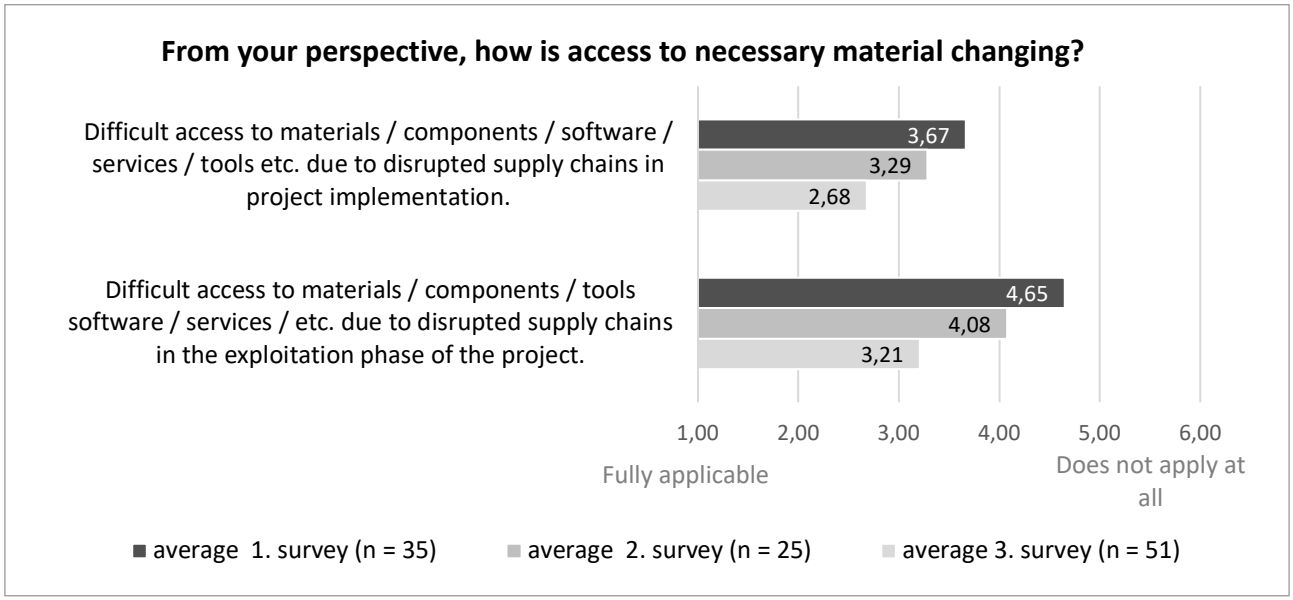




4.2 PROGRAMME B

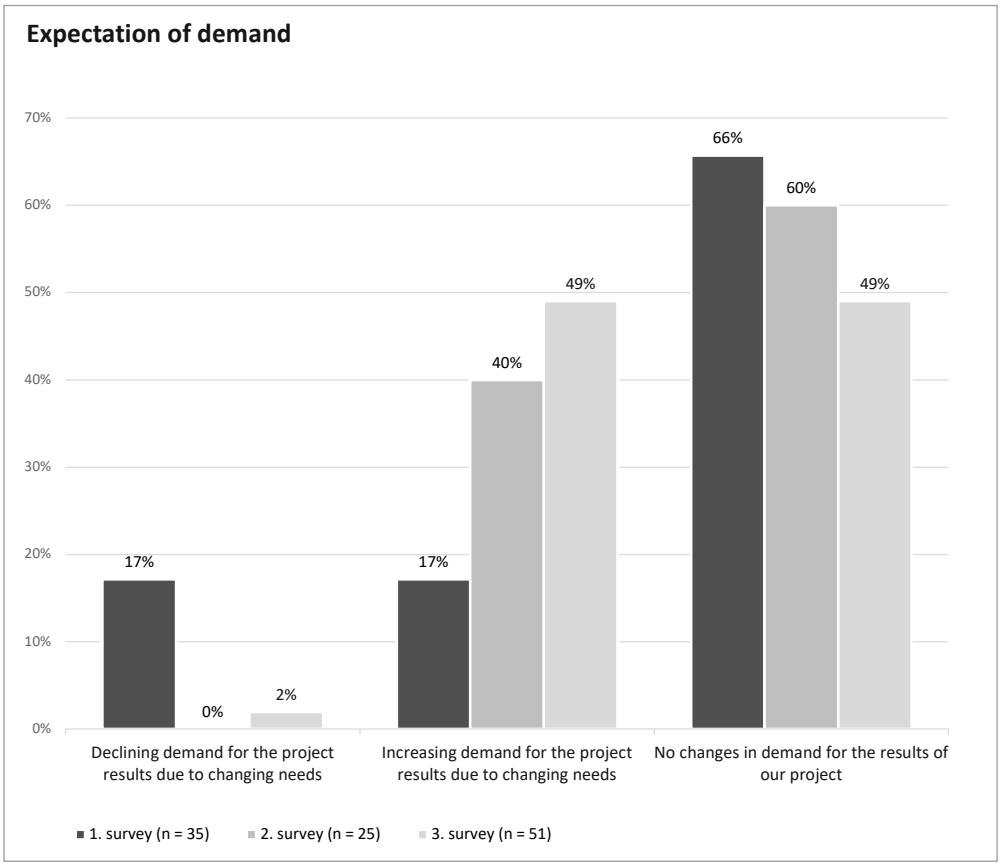
In the evaluation of programme B, data from three survey waves of projects still running between summer 2020 and summer 2021 were analysed. It can be seen that the pandemic affected the logistics chains, which led to significantly more difficult access to material in this programme. This effect increased over time (Figure 4).

Figure 4: Access to material. 1st-3rd survey in the period summer 2020 to 2021



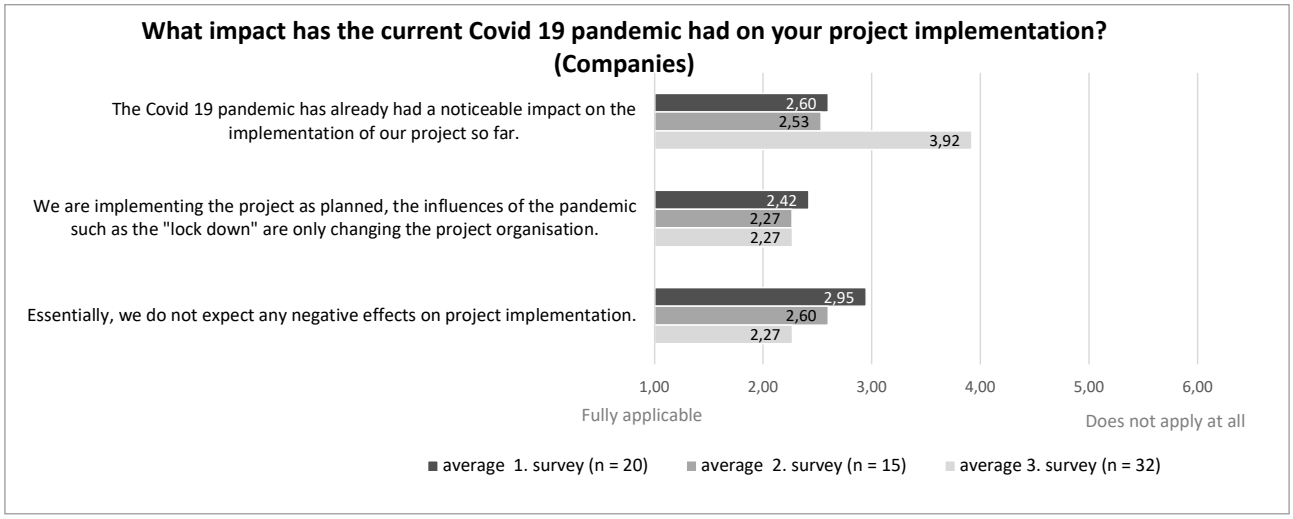
At the same time, a change in demand for the project results is expected in the course of the survey waves in this programme. While this effect was predicted to be negative or positive to a similar extent in the first survey (summer 2020), the expectation of a positive effect clearly prevails in the latter survey (summer 2021).

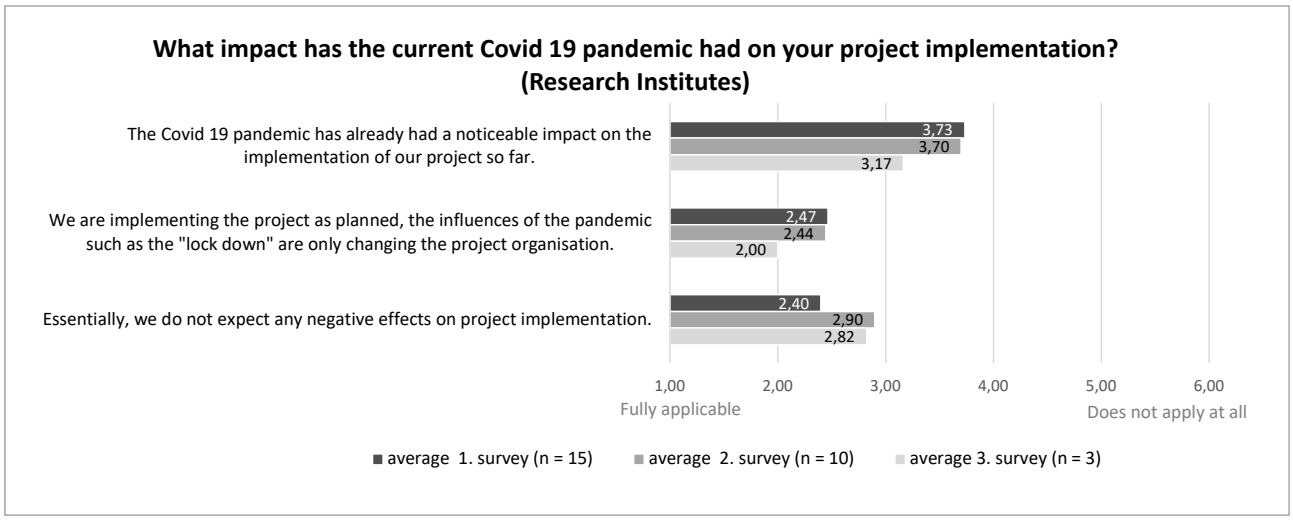
Figure 5: Expectation of demand for project results. 1st-3rd survey in the period summer 2020 to 2021



Since Programme B involves the funding of collaborative projects, different effects can be analysed here according to the type of actor. While in the first survey wave in summer 2020 companies reported more clearly on the possible effects of the pandemic ("noticeable impact") and are more optimistic in summer 2021 ("we do not expect any negative effects"), research institutions initially assessed the effects of the pandemic more cautiously but are more critical in summer 2021 (Figure 6).

Figure 6: Impact on project implementation (companies vs. research institutions)



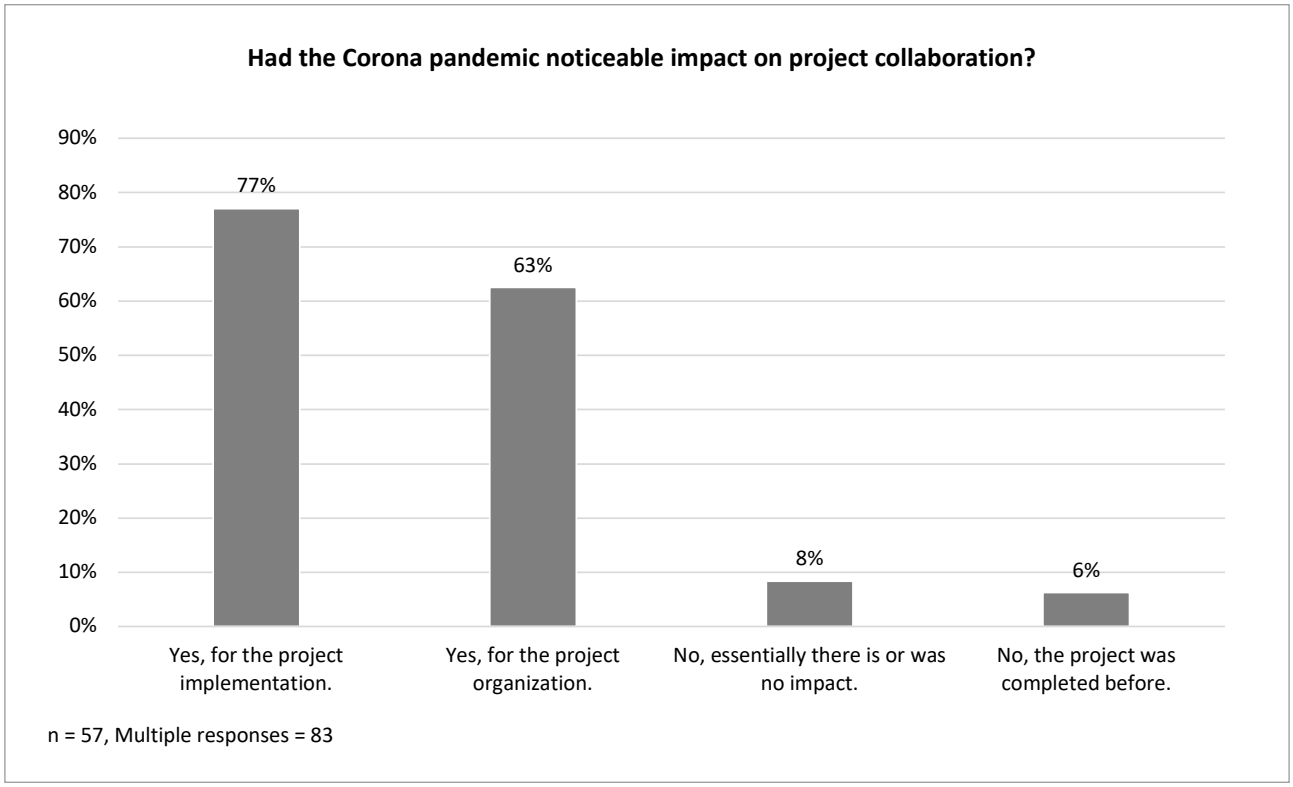


Since Programme B involves the funding of collaborative projects, different effects can be analysed here according to the type of actor. While in the first survey wave in summer 2020 companies reported more clearly on the possible effects of the pandemic ("noticeable impact") and are more optimistic in summer 2021 ("we do not expect any negative effects"), research institutions initially assessed the effects of the pandemic more cautiously but are more critical in summer 2021 (Figure 6).

4.3 PROGRAMME C

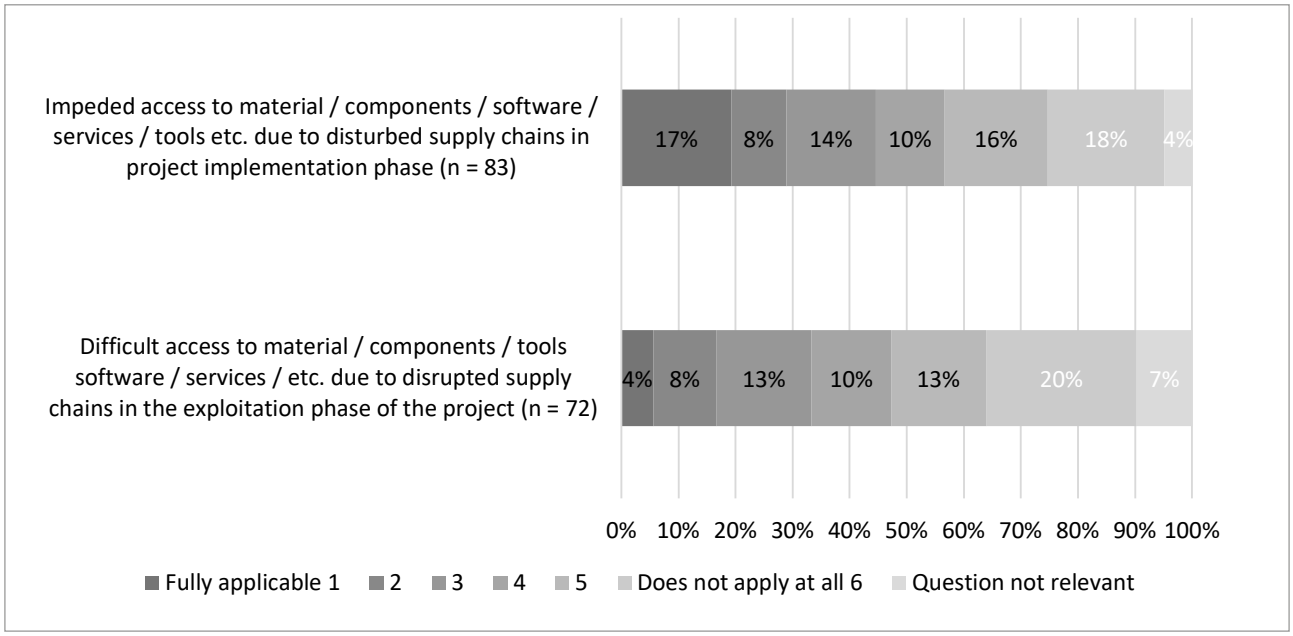
In Programme C, a good two-thirds of respondents reported strong effects on project implementation (77 percent) as well as project organisation (63 percent) (Figure 7).

Figure 7: Impact of the Corona pandemic on the joint project



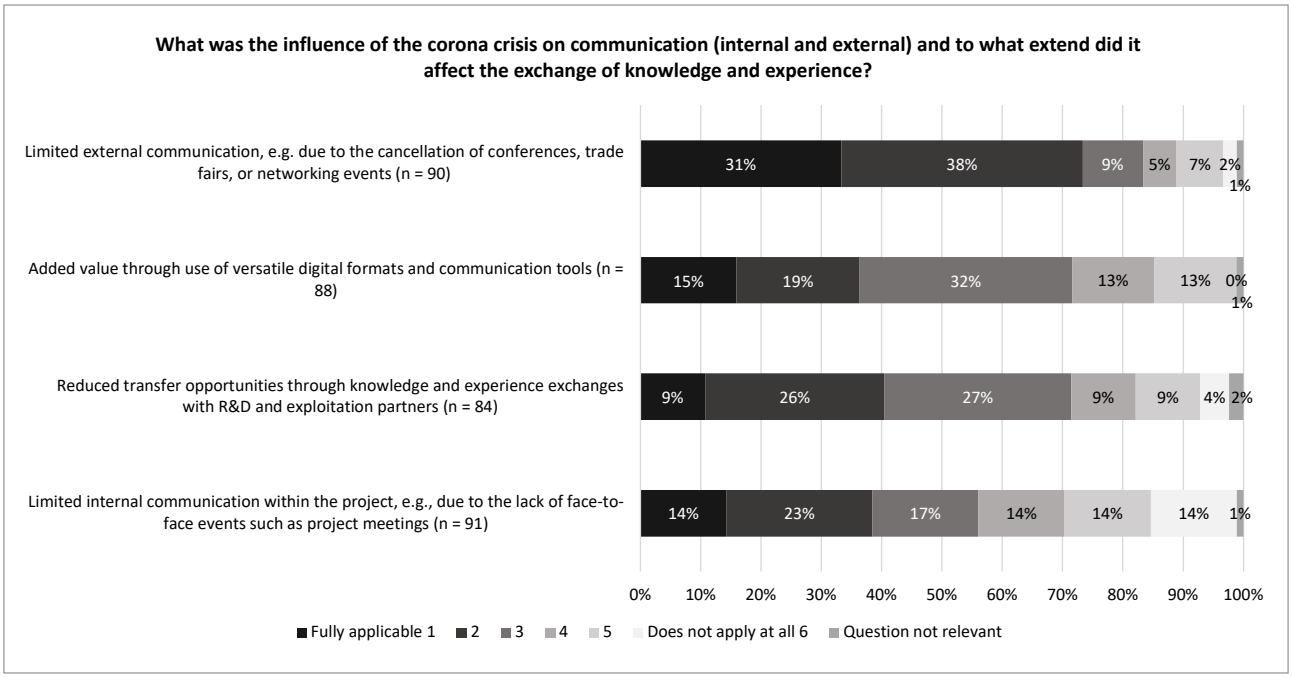
In contrast, the grantees' perception of the Corona effect on access to materials, components, software, etc. is mixed. Respondents vary in the extent to which they are affected by the Corona-related effects, from high effects to no effects at all (Figure 8).

Figure 8: Impact of the Corona pandemic on project implementation
Source: Online survey, representation iit



The survey also aimed to explore potential impacts on communication (Figure 9). Limited external communication was perceived as a key effect of the Corona pandemic. More than two-thirds of the grantees (69 percent) stated that, for example, due to the cancellation of conferences, trade fairs or networking events, the exchange on project-specific and cross-cutting issues suffered. This gap was compensated for by digital formats and communication tools. One third (34 percent) of the respondents see added value here. Another relevant hurdle is seen as limited internal communication, because face-to-face events such as project meetings have been eliminated. About one third (37 percent) of respondents agree that internal communication has deteriorated, although almost one third (28 percent) also see no perceptible deterioration in internal communication. Finally, about one third (35 percent) of respondents perceive a deterioration in transfer opportunities due to reduced exchange with R&D and exploitation partners.

Figure 9: Impact of the Corona pandemic on communication



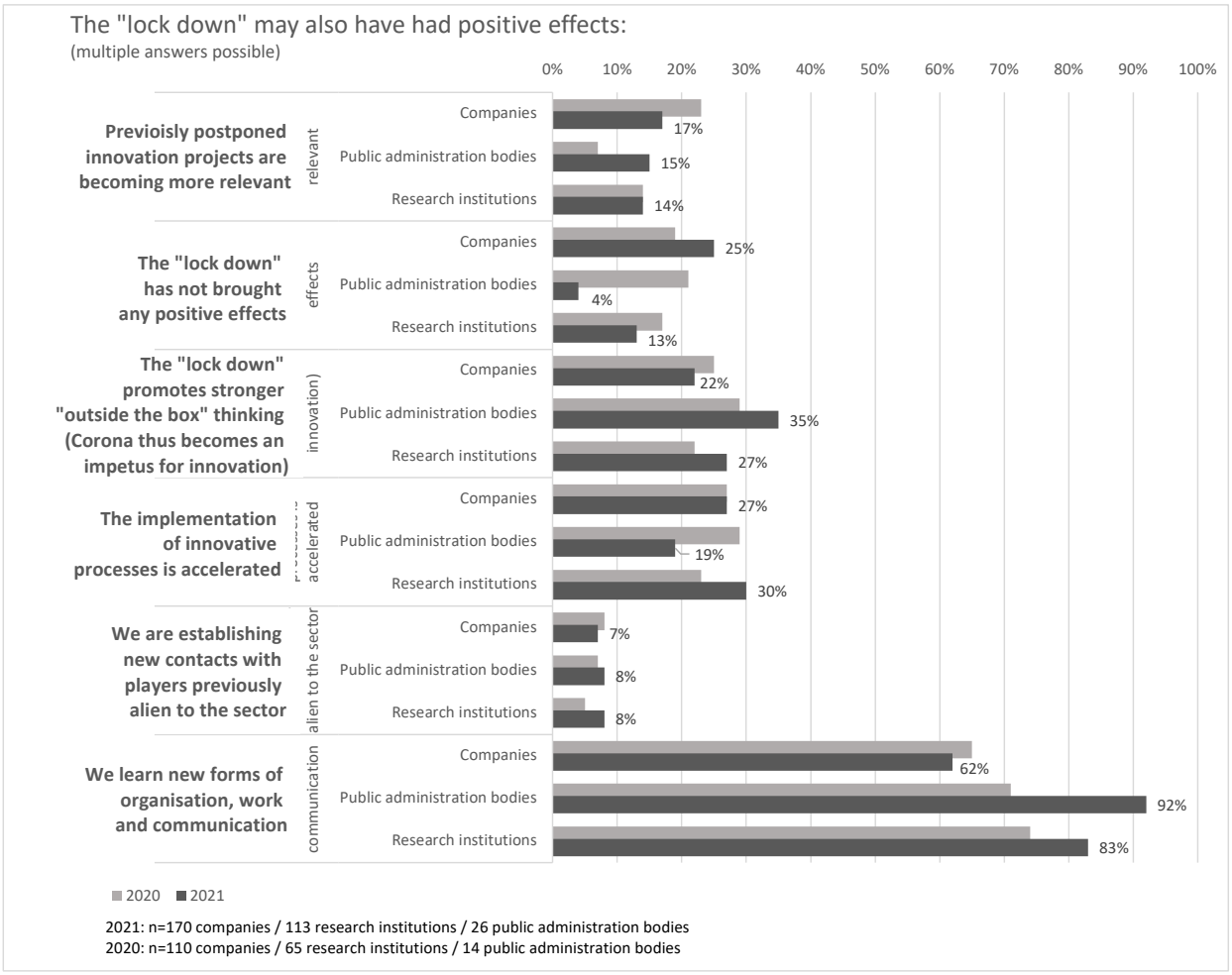
In programme C, respondents also provided detailed information on the obstacles that led to delays. For example, planned projects were cancelled during the pandemic (e.g., final demonstration, joint meeting for a project conclusion, cancellation of a study with test persons). Supply bottlenecks also meant that projects could not be implemented as planned. Communication was sometimes severely restricted, e.g., due to home office arrangements and short-time work during the lockdown. This resulted in considerable time delays. The change in user behaviour had an impact on the success of individual projects because relevant project goals for researching user acceptance could not be implemented to the planned extent. Access to laboratories as well as business trips to project partners also became more difficult for some project participants.

4.4 PROGRAMME D

In Programme D, funding is provided not only to companies and research institutions, but also to municipal agencies and authorities. Here, the survey data was evaluated at the actor level (companies, research institutions, municipalities and administrations). One of the questions asked was whether there were also possible positive effects from the pandemic (Figure 10). It turns out that the crisis has had no positive effect at all for only a few respondents. In fact, in 2020, around two-thirds of respondents already reported positive experiences with regard to new forms of organisation, work and communication. In 2021, this effect is perceived by the research institutions and municipalities/public administration institutions to be significantly stronger than in the previous year. In contrast, only a minority report positive effects that go beyond this, whereby differences between the groups of actors can also be observed here. In 2021, one third of the administrations stated that the "lock down" had led to increased "out of the box" thinking within their institutions. This effect was less pronounced in the case of companies. For every fourth company, the "lock down" led to an acceleration of innovative processes.

In principle, the data do not indicate any major differences between the situation in the summer of 2020 and the summer of 2021; the prevailing positive effects are constant across all groups of actors.

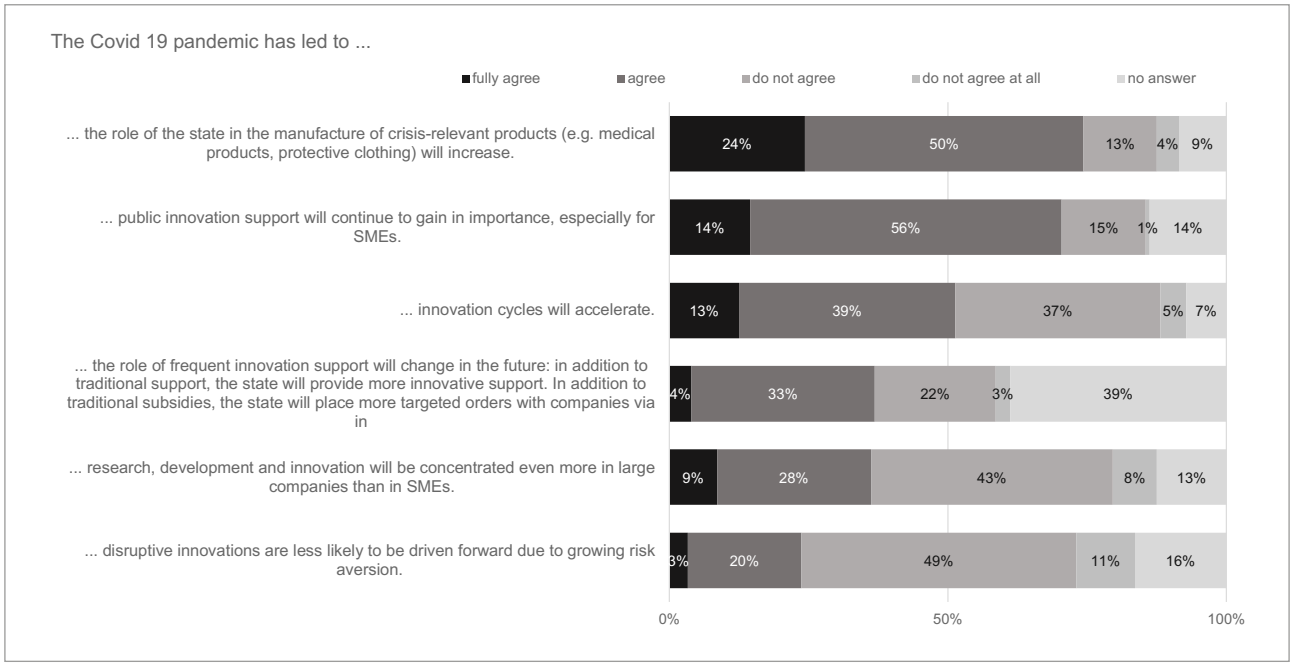
Figure 10: Positive effects from the lock down



4.5 PROGRAMME E

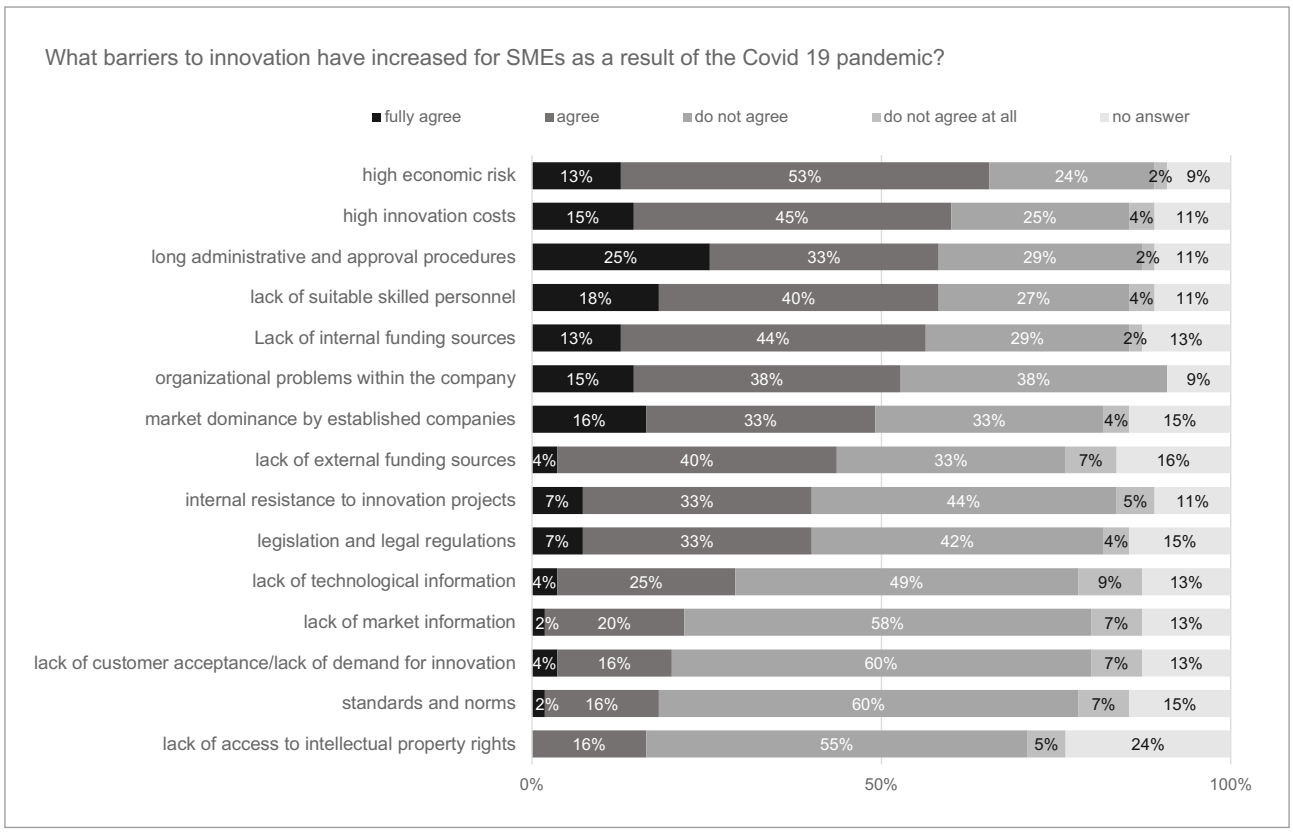
In Programme E, 152 experts were asked about the impact of the Corona pandemic on research and innovation. A large proportion agreed that innovation cycles accelerated in the wake of the pandemic and that the importance of public funding increased at the same time (Figure 11). In the future, a stronger role of the state in crisis management is seen, for example, in the production of protective clothing (74 percent agreement) and in innovative public procurement for the targeted promotion of companies (59 percent agreement). In the view of the experts surveyed, innovation promotion will become more important, especially for SMEs (70 percent agreement). This is also considered important because the respondents assume that R&D efforts will focus even more on large companies in the future, while SMEs could fall behind (80 percent agreement). When asked whether a possible risk aversion could lead to fewer disruptive innovations in the future, only one fifth of the respondents are of this opinion (23 percent agree).

Figure 11: Positive effects from the lock down



The survey results also indicate that the Corona pandemic causes barriers to innovation, especially for SMEs (Figure 12). The top three barriers to innovation are cited as: economic risk (66 percent agreement), innovation costs (60 percent agreement) and long administrative and approval procedures (58 percent agreement).

Figure 12: Innovation barriers in the wake of the Corona pandemic for SMEs



Other barriers to innovation that have increased as a result of the Corona pandemic in the view of the experts interviewed are access to skilled personnel, financing and market dominance by individual, established companies; internal organisational barriers are also mentioned.

5 CONCLUSIONS

This contribution aimed to provide initial explanations of the Corona pandemic's effects on innovation promotion and chosen adaptation strategies. In addition, the question of how the "Corona crisis" can be adequately captured in evaluations was explored. Our investigations show that the impact assumptions we made in the impact model can largely be traced with Covid19 as an external influencing factor.

Particularly at the beginning of the pandemic, the effects led to a reduction in internal resources, restrictions in staff availability, as well as changes in priorities and resulting adjustments in the timing and content of research projects.

Supply chains were interrupted and access to materials became more difficult. External communication in particular deteriorated because participation in trade fairs and conferences was not possible. The contact restrictions also had a significant impact on internal communication in the project teams. Especially for projects that started or were still running during the pandemic, there was a risk of longer running times and difficulties in marketing the project results in a timely manner. Project organisation and implementation was significantly more difficult.

Projects that have already been completed for a longer time and are already in the transfer phase were somewhat less affected by the corona crisis. Also, not all types of actors were affected to the same extent; while some suffered very great disadvantages, others were only confronted with a few restrictions. Differentiated by actor groups, the effects of the pandemic were felt particularly strong in municipalities and public administration, while companies were the least affected. The public sector revealed its significant digitalisation backlog. For example, it found it much more difficult to switch its communication channels to video conferencing or to provide laptops for mobile working.

Many of the negative impacts in the research projects cannot be compensated for in the project period. Overall, this can lead to a somewhat poorer programme success. This special negative effect in the programme impact must be taken into account in future comparisons of funding programmes.

The survey results also show that expectations of demand for their project results changed over time. While many of the respondents were still very pessimistic about the demand for their project results in the summer of 2020, they were already more optimistic in the summer of 2021. This was also reflected in the first signs that the economy would develop positively again from 2022 at the latest.

Greater importance is being attached to the state as a stimulus and promoter. Innovative public procurement and innovation support, especially for SMEs, are seen as important levers. As outlined in the impact model, such Corona-induced political measures can have a positive effect on the results of projects and thus on the overall programme.

Last but not least, positive effects could also be traced, such as the digitalisation boost and an increase in innovation activity. New communication formats were tested, and innovations promoted through outside-the-box thinking. During the crisis, companies and employees experienced how quickly problems can be solved and changes can be shaped.

The data available so far show that the effects change over time. For this reason, an analysis of further effects in the coming years would be interesting in order to be able to estimate medium-term effects. This also concerns structural effects, e.g., in the internal organisation of research projects on the part of project participants. At the moment it seems that the effects point in different directions and will be difficult to quantify - depending on the programme, there may have been an increase or decrease of demand or a delay in demand due to delivery difficulties, innovation may have accelerated or slowed down. Comparisons between programmes (e.g., in meta-evaluations) will need to consider this in detail as one cannot assume that the economic environment has continued to develop equally for all sectors and actors.

Also still outstanding is a systematic analysis of the adjustment steps on the part of the programme managers. In the short term, the requirements for project participants in the pandemic were relaxed in some funding programmes in order to mitigate specific challenges. Likewise, project extensions were approved more generously. Whether these adjustments to the programmes and gains in flexibility will last or be reversed is still an open question.

It also became apparent that innovation cycles have accelerated in the wake of the pandemic. The digitalisation boost is influencing and accelerating the transformation of important policy fields (climate/sustainability, mobility, health). Underlying reasons - perhaps because everything is in upheaval anyway (new actors, geopolitical changes) and the questioning of established structures and processes by the pandemic favour this - need to be investigated.

The pandemic has also shown that many things can be done more efficiently and less expensively. This learning process is just in its infancy. An intensive discussion about more agility and new funding formats has begun. More courage to experiment with new formats is desirable, ideally accompanied by evaluations, to see whether the cost-benefit ratio changes positively in the long term.

6 LITERATURE AND SOURCES

Azagra-Caroa, Joaquín M. / Tijssen, Robert J.W. / Turde, Elena M. / Yegros-Yegros, Alfredo (2020): University-industry scientific production and the Great Recession. https://www.sciencedirect.com/science/article/pii/S0040162517304067?via%3Dihu

Dachs, Bernhard / Peters, Bettina (2020): Covid-19-Krise und die erwarteten Auswirkungen auf F&E in Unternehmen. https://www.zew.de/publikationen/covid-19-krise-und-die-erwarteten-auswirkungen-auf-fe-in-unternehmen

Deutscher Bundestag (2020): Antwort der Bundesregierung auf die Kleine Anfrage der Abgeordneten Dr. Anna Christmann, Kai Gehring, Dieter Janecek, weiterer Abgeordneter und der Fraktion BÜNDNIS 90/DIE GRÜNEN – Drucksache 19/21074 – Zukunftsinvestitionen und innovationspolitische Maßnahmen im Konjunkturpaket der Bundesregierung. https://dserver.bundestag.de/btd/19/218/1921858.pdf

DFG (2021): DFG Infobrief. Zahlen und Fakten zur Forschungsförderung (https://www.dfg.de/download/pdf/dfg_im_profil/geschaeftsstelle/publikationen/infobriefe/ib01_2021.pdf)

Economist (4.4.2020): Big tech's covid-19 opportunity. https://www.economist.com/leaders/2020/04/04/big-techs-covid19-opportunity

Europäisches Patentamt (16.3.2021): Healthcare innovation main driver of European patent applications in 2020. https://www.epo.org/news-events/press/releases/archive/2021/20210316.html EPA 3/2021

Estermann, Thomas (2020): University funding amid the coronavirus pandemic: lessons learnt from the 2008 global economic crisis. https://www.eua.eu/resources/expert-voices/165-university-funding-amid-the-coronavirus-pandemic-lessons-learntfrom-the-2008-global-economic-crisis.html?utm_source=social&utm_medium=Twitter&utm_name=Twitter-social-23-4-2020

EY (4/2021): Startup-Barometer Europa 2021. https://www.ey.com/de_de/news/2021/04/ey-startup-barometer-europa-2021

KfW (2021): KfW-Gründungsmonitor 2021. https://www.kfw.de/PDF/Download-Center/Konzernthemen/Research/PDF-Dokumente-Gr%C3%BCndungsmonitor/KfW-Gr%C3%BCndungsmonitor-2021.pdf

Munich Startup (26.2.2021): EXIST: Neuer Rekord bei Anträgen auf Gründerstipendium. https://www.munich-startup.de/69327/exist-neuer-rekord-bei-antraegen-auf-gruenderstipendium/

Rammer, Christian / Schubert Torben (2016): Concentration on the Few? R&D and Innovation in German Firms 2001 to 2013. http://ftp.zew.de/pub/zew-docs/dp/dp16005.pdf

Kerlen, Christiane et. al. (2020): Corona als externer Schock für Innovationsförderprogramme - erste Ergebnisse laufender Evaluationen und Monitoringprozesse. https://www.iit-berlin.de/publikation/corona-als-externer-schock-fuer-innovationsfoerderprogramme-erste-ergebnisse-laufender-evaluationen-und-monitoringprozesse/

ZEW (11/2020): Junge Unternehmen in der Corona-Krise im Frühjahr 2020. http://ftp.zew.de/pub/zew-docs/gutachten/JungeUnternehmen-CoronaKrise2020.pdf

ZEW (2/2021): Innovationen in der deutschen Wirtschaft. Indikatorenbericht zur Innovationserhebung 2020. https://www.bmbf.de/upload_filestore/files_202102xx/MIP_2020_05.pdf

ZEW (7/2021): The Impact of the Covid-19 Crisis on Innovation: First Insights from the German Business Sector. https://www.zew.de/fileadmin/FTP/ZEWKurzexpertisen/EN/ZEW_Shortreport2106.pdf?v=1625757219

AUTHORS

CHRISTIANE KERLEN
Kerlen Evaluation
16 West Savile Gardens, Edinburgh EH9 2AB
E: christiane.kerlen@kerlen-evaluation.de

SONJA KIND
iit – Institut für Innovation und Technik c/o VDI/VDE-IT
Steinplatz 1
10623 Berlin
Germany

KAROLINE RODRIGUEZ
iit – Institut für Innovation und Technik c/o VDI/VDE-IT
Steinplatz 1
10623 Berlin
Germany

LEO WANGLER
iit – Institut für Innovation und Technik c/o VDI/VDE-IT
Steinplatz 1
10623 Berlin
Germany

GUIDO ZINKE
iit – Institut für Innovation und Technik c/o VDI/VDE-IT
Steinplatz 1
10623 Berlin
Germany

JAN WESSELS
iit – Institut für Innovation und Technik c/o VDI/VDE-IT
Steinplatz 1
10623 Berlin
Germany
E: wessels@iit-berlin.de

KEYWORDS: R&D Funding; Evaluation, Covid 19

TOWARDS A FRAMEWORK FOR IMPACT ASSESSMENT FOR MISSION-ORIENTED INNOVATION POLICIES. A FORMATIVE TOOLBOX APPROACH

FLORIAN WITTMANN, FLORIAN ROTH, MIRIAM HUFNAGL, RALF LINDNER AND MERVE YORULMAZ

DOI: 10.22163/fteval.2022.540

Mission-oriented policies (MOIP) have become important means to foster transformative change in many countries. Yet, approaches for assessing these policies' impacts are still in their infancy, not least due to the complexity of MOIP. To address this gap, we propose a toolbox approach that supports policy-makers during policy design and implementation, and allows for an identification of potential impacts by a theory-based approach. To disentangle the complexity of missions, we first conceptualize MOIPs as multiple translation processes from mission formulation and design to implementation. Each translation step shapes the policies' impacts. Based on this framework, we develop a set of specific analytical tools that are intended to support the process of bringing missions into realization, but also help to assess whether missions contribute to the postulated goals. These tools include a mapping of the socio-technical systems, a typology to explore the transformative ambition of missions, a process to develop impact pathways, an inventory of policy instruments to support the mission design, and indicators to measure mission progress along the developed pathways. Finally, we propose several analytical questions to explore the context for the development of potential impacts.

INTRODUCTION

In recent years, mission-oriented innovation policies (MOIP) aiming at transforming socio-technical systems have gained increasing attention in both academic debates (Larrue 2021; Mazzucato 2017; Robinson et al. 2019) and among policy-makers, as can be seen in numerous initiatives at different levels (German High-Tech Strategy, missions in Horizon Europe, etc.). The promise of catalyzing transformative change through coordinated cross-sectoral action, actor mobilization and a stronger directionality of science, technology and innovation (STI) policies, has gained momentum against the background of the challenges societies are facing. Ideally, MOIP apply a variety of policy instruments that *'span different stages of the innovation cycle, from research to demonstration and market deployment, mix supply-push and demand-pull instruments, and cut across various policy fields, sectors and disciplines'* (Larrue 2021, p. 11).

With the growing number of MOIP initiatives, questions about the implementation and the ability to monitor and evaluate such approaches have come to the forefront (Dinges et al. 2020; Janssen et al. 2021; Larrue 2021; Weber et al. 2014). However, the analysis of impacts involves multiple conceptual challenges, including the multidimensionality and interaction of effects, the different analytical levels, the long time horizon associated with mission goals and the empirical diversity of missions to be found under the MOIP label (cf. e.g. Amanatidou et al. 2014; Arnold et al. 2018; Edler et al. 2012; Kuittinen et al. 2018; Magro et al. 2019; Weber et al. 2014; Wittmann et al. 2021a; Wittmann et al. 2021c). So far, there have been first attempts to evaluate and assess the impact of individual programs and strategies with mission-orientation from program or innovation systems perspectives (Bührer et al. 2020; Hekkert et al. 2020; Hüsing et al. 2017; Wesseling et al. 2020), as well as general frameworks for MOIP (Weber et al. 2014, p. 9) and transformative innovation policies in complex settings (Arnold et al. 2018; Ghosh et al. 2021; Grillitsch et al. 2019; Janssen 2016).

In this contribution, we propose a flexible and formative toolbox approach that enables evaluators to investigate the potential effects of missions, but at the same time supports policy-makers to formulate, design and implement MOIP successfully.[1] The toolbox draws on identified requirements for the evaluation and impact assessment of MOIP and transformative policies in general, offering a diverse set of analytical elements that can be employed selectively or in combination, depending on the specific conditions and contexts.

3 A more comprehensive description can be found in the final report of the scientific support action to the German High-Tech Strategy 2025: https://www.isi.fraunhofer.de/content/dam/isi/dokumente/ccp/2021/Endbericht_Wirkungsmessung_englisch.pdf

KEY REQUIREMENTS FOR ASSESSING THE IMPACT OF MOIP

In recent years, the debate around the challenges of impact assessment and evaluation of innovation policies has increasingly shifted to the requirements for assessing transformative innovation policies such as MOIP (Molas-Gallart et al. 2021; Wittmann et al. 2021c). Drawing on these insights, we postulate four key requirements to better understand MOIP and offer systemic guidance to the policy process.

1) A STRONG FORMATIVE PERSPECTIVE PROVIDING PRACTICAL GUIDANCE

The turn towards cross-cutting and transformative change has entailed calls for a formative perspective (Magro et al. 2019; Molas-Gallart et al. 2021). Imposing a long-term perspective and considerable requirements with regard to cooperation and coordination across different fields, actors, etc., a framework for impact assessment should effectively support policy actors during the implementation of this policy approach and provide the opportunity for feedback and learning. Thereby, it supports reflecting the experimental and dynamic character of missions. For this purpose, the emphasis is put on an approach that provides practical guidance making research insights on MOIP usable for implementation, as the practical realization of missions continues to be a considerable challenge for public actors colliding with established routines and institutional arrangements (Lindner et al. 2021). This also includes a shift towards an increased reliance on ex-ante elements to inform the process, as postulated by Weber and Polt (2014).

2) A COMPREHENSIVE AND INTEGRATED PERSPECTIVE GRASPING ALL PHASES OF MOIP

Closely related to the aforementioned point is an integration of the framework into the realization process of missions. Therefore, we need to take into consideration all phases of MOIP, including mission (policy) formulation, design and implementation. Previous research has demonstrated that the formulation process of mission goals can be considered as crucial for later success in implementation (Janssen et al. 2020; Lindner et al. 2021) as mission realization can be considered to consist of different linked phases (Wittmann et al. 2021b) that usually emerge in the context of existing policy traditions and fields (Larrue 2021). Therefore, we need to acknowledge the very specific complex negotiation processes at different levels that are associated with this approach. To capture this complexity, we draw on Kroll (2019) and the concept of translation processes for evaluation and its application to MOIP (Wittmann et al. 2021b) Applying this perspective, we conceptualize the process of carrying out missions as multiple interconnected translation processes at different levels that shape and constrain the ability of missions to realize impacts:

- **Mission formulation:** The translation process is necessary to narrow down the mission towards a specific goal. This process is dependent on the specific context (Edler et al. 2020) and may lead to different interpretations on how to achieve these goals (Wittmann et al. 2021a). This decision at the strategic level has profound implications for the later stages as the legitimacy and urgency of missions affect the ability to mobilize actors and resources (cf. Janssen et al. 2020; Larrue 2021)
- **Mission design:** The second translation occurs at the stage of policy-makers designing the activities that are encompassed by a mission through a deliberate choice of inputs. To translate goals into specific measures and instruments it is necessary to combine different types and generations of policy instruments, which might lead to policy-layering but also include newly designed instruments.
- **Mission implementation:** The final translation relates to the step from mission design to mission implementation, focusing on administrations and funding agencies bringing instruments into realization. These implementation activities of actual instruments are the prerequisite for the unfolding of the intended effects of a mission in the long run.

3) A THEORY-BASED AND PROCESS-ORIENTED APPROACH TO STUDY IMPACTS

One key complication of the analysis is the fact that missions require to take both the project- and systemic-level into consideration (Amanatidou et al. 2014; Weber et al. 2014). Theory-based evaluations are commonly considered as a useful tool for the evaluation of complex policies, as they are able to contrast actual developments with previously derived expectations (Arnold et al. 2018; Arnold 2019; Belcher et al. 2020; Bührer et al. 2019; Joly et al. 2015; Joly et al. 2017; Joly et al. 2019; Kalpazidou Schmidt et al. 2017a; Miedzinski et al. 2013; Molas-Gallart et al. 2021). This provides the opportunity for tracking the progress of missions even when effects are more systemic or are only expected to materialize in the long run and cannot be controlled by the mission owners (cf. Belcher et al. 2020). Moreover, this approach has proven to be appropriate for a context-sensitive perspective, accounting for the fact that dynamics may play out differently (Kalpazidou Schmidt et al. 2017b) depending on the topic, institutional context and existing policies (cf. Wittmann et al. 2021b). At the same time, missions build heavily on a different approach to policy-making being in contrast to established practices (Lindner et al. 2021), implying that an input-output perspective will not be sufficient to explore whether appropriate conditions/the existence of hindering factors for the materialization of effects are in place. Consequently, we propose to combine a theory-based approach with impact pathways and a process-oriented analysis of the translation processes (see above) that can be considered as a key bottleneck of such policies.

4) A FLEXIBLE AND MODULAR APPROACH

Given the considerable diversity of activities that can be observed in the context of MOIP (Griniece et al. 2018; e.g. Kuittinen et al. 2018; Larrue 2021; Polt et al. 2019; Wanzenböck et al. 2020; Wittmann et al. 2021a), a mission exhibits a highly specific profile with regard to the scope, the domain, and the way changes are to be achieved and the role of different types of instruments. In consequence, missions are highly di-

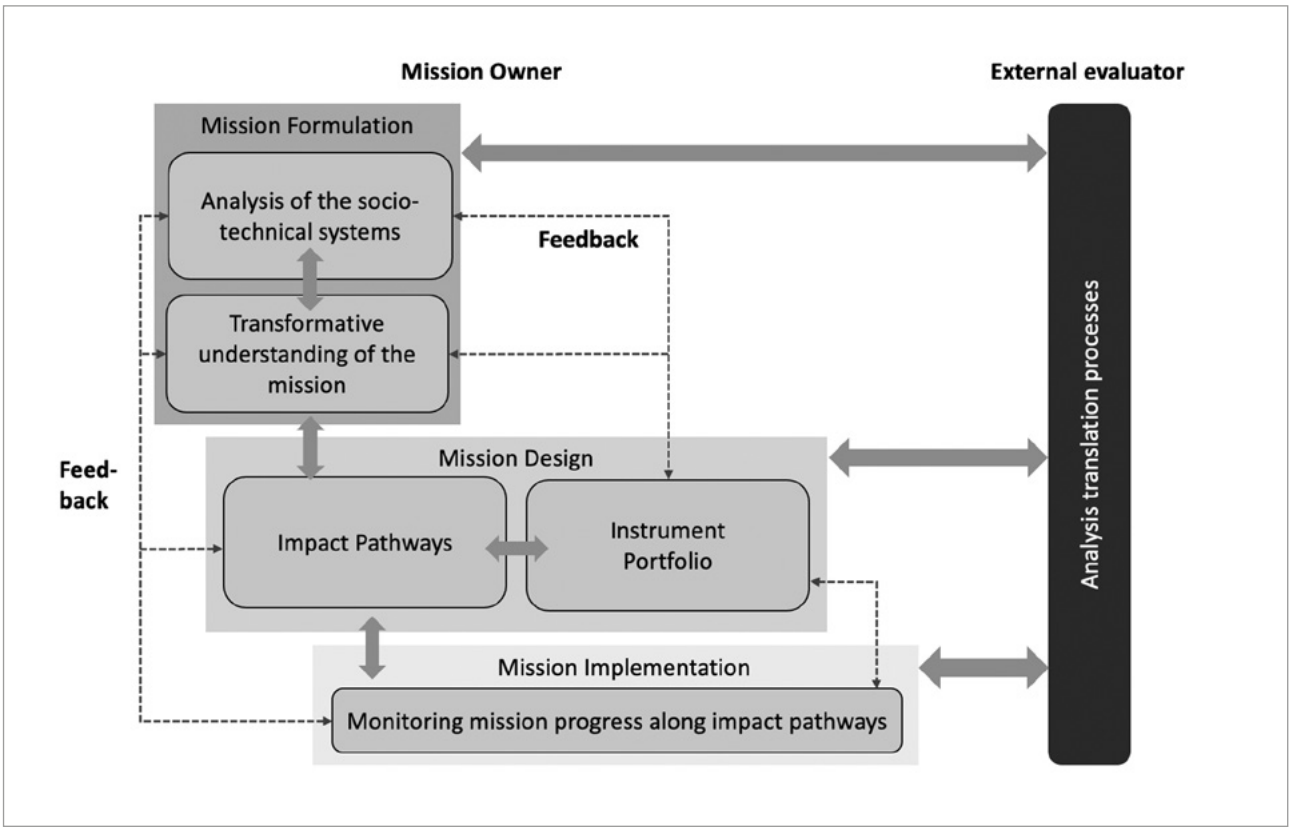
verse so that there is no blueprint how different areas (science, economy, society) will interact with each other and how important their role is relative to each other. For this reason, departing from the idea that missions in general aim for an overarching societal impact, a flexible and modular approach is indicated. Providing a set of stylized types of missions, ways of intervention etc., these can be applied to actual missions in order to develop a context-based framework for mission evaluation. Emphasizing the importance of a modular and flexible approach that fits the specific context also implies methodological openness and accounting for different ways of stakeholder involvement. In consequence, the framework does not aim to prescribe the use of a certain method, as the appropriateness may be conditional on the context and available resources.

A TOOLBOX APPROACH FOR ASSESSING MISSION-ORIENTED POLICIES

We propose a modular approach that addresses the different levels of mission-oriented policies and thereby fulfills the aforementioned requirements. For this purpose, we introduce a total of six closely connected toolbox elements that can support the implementation process through creating awareness among stakeholders and provide learning and feedback opportunities (see Figure 1). Five of these elements are directly linked to mission owners, i.e. those carrying out the missions. These elements may stimulate self-reflection processes among mission owners, provide guidance on decision-making and prepare the ground for a systematic impact assessment by putting the necessary features in place. These elements are complemented by a cross-cutting analysis of the translation processes of missions that constitute the key pillar for evaluation putting mission activities into context. The translation processes shape the realization of missions and provide the context for the manifestation of impacts. In sum, we consider the impact assessment as an integral part of a mission that needs to be closely aligned with the main elements of the mission from the very beginning.

As indicated by the feedback loops, we do not see the process as linear but as iterative, supporting learning effects by involved stakeholders between different elements. In practice, this implies that activities may temporally overlap, as e.g. the development of appropriate impact pathways and the identification of suitable instruments might affect each other. While emphasizing the importance of combining the different elements, the approach is not deterministic about methods and only describes the overall frame. Thereby, it acknowledges the potential existence of varying understandings, resources and priorities such as the extent to which missions aim for the integration of stakeholder involvement – a decision that is ultimately dependent on the mission owner.

Figure 1: Overview of the elements of the toolbox for assessing MOIP (own elaboration)



1) ANALYSIS OF THE SOCIO-TECHNICAL SYSTEM

A comprehensive picture of socio-technical systems can contribute to a better understanding of key societal challenges, but also of transformative policies that aim to alter the configuration of these systems. Socio-technical systems can be understood as the 'articulated ensembles of social and technical elements which interact with each other in distinct ways, are distinguishable from their environment, have developed specific forms of collective knowledge production, knowledge utilization and innovation, and which are oriented towards specific purposes in society and economy' (Borrás et al. 2014, p. 11). A detailed understanding of system complexities is needed to inform public (policy makers) as well as private actors (industry, consumer) about their role, responsibility and agency for systemic change.

A system mapping process presents an illustrative technique that serves a double purpose: a) as an analytical tool to (collaboratively) map out complex topical landscapes (looking back: capturing the status quo), b) as an explorative strategic tool to gather important topics, policies, actors, and system boundaries that need to be engaged before policy formulation starts (looking forward: depicting future needs). System mapping is a promising approach for better understanding complex challenges, but also to guide the development of solutions (Cavill et al. 2020; Matti et al. 2020). Employing this eagle's perspective on complex socio-technical systems overcomes several shortcomings of classic tools of policy analyses, because it starts off with a clear delineation of the major topics and subtopics, providing the analytical basis for further investigation. Moreover, the association of key players with concrete policies or other measures can help to better understand the complex interdependencies between system elements.

Mapping the socio-technical system that is to be transformed may prove particularly useful in the process of mission formulation by the mission owners. At this stage of the mission process, it can enhance the mission owners' understanding of key challenges, but also supports actors in clarifying the boundaries of the system a mission aims to transform. Creating awareness and consensus on these question can facilitate the discussion about the problem-solution space that characterizes MOIP (Wanzenböck et al. 2020). Moreover, grasping the overall complexity of socio-technical systems that relate to the challenges at stake provides a baseline for the subsequent assessment of translation processes.

2) EXPLORING THE TRANSFORMATIVE AMBITION

While assuming that all missions pursue a transformative agenda, there exist different understandings of how to achieve the desired changes. This is also reflected in the growing diversity of empirical missions and the academic attempts to conceptualize variations between missions (Polt et al. 2019; Wanzenböck et al. 2020; Wittmann et al. 2021a). Whereas some missions emphasize the role of technological/scientific innovation, others explicitly aim for behavioral changes as part of the transformative agenda. Exploring the transformative understanding of a mission can support the mission owners by pinpointing at requirements and consequences of these decisions and providing guidance for the process of mission design. At the same time, it prepares the ground for the ex-ante assessment of mission design, trying to understand whether activities in the mission context are compatible with the postulated goals. Whereas transformer missions are likely to require a more comprehensive instrument cross-cutting different fields, mission resembling an accelerator type will be sufficient with a narrower focus. In the following, we apply the typology developed by Wittmann et al. (2021a) that distinguishes between four types of missions characterized by specific challenges during implementation (see Table 1) that can serve as a point of reference for mission owners when deciding about the scope and character of their mission formulation.

Table 1: Different types of missions and key features (based on Wittmann et al. 2021a)

	Accelerator Mission		Transformer Missions	
	Type 1 (A1)	**Type 2 (A2)**	**Type 1 (T1)**	**Type 2 (T2)**
Motivation	Problem-solving	Solution-driven	Solution-driven	Problem-solving
Main logic of change	Scientific/ technological change	Bringing knowledge to application	Reconfiguration of sectoral logics	System transformation (incl. behavioral change)
Key stakeholders	Science	Science, Economy	Science, Economy, collective sectoral actors	Science, Economy, collective sectoral actors, civil society
Instrument mix	Mainly STI (distribution)	Mainly STI (distribution, systemic management)	Broad (distribution, regulation, information)	Broad (re-distribution, regulation, information)
Coordination requirements	Limited	Medium	High	Very high
Main challenges	Uncertainty, long-time horizons, shared understanding of problem, achieving critical mass for change	Ensuring appropriate framework conditions, overcoming existing bottlenecks, achieving critical mass for change	Dealing with path-dependencies/lock-ins, integration of sectoral policies, shift towards systemic change	Re-distribution/ compensating potential losers, involving society & different levels, shift towards systemic change

3) DEVELOPING IMPACT PATHWAYS

Based on the mission goals formulated, the next step is the design of the mission. A first key element in this regard is the development of appropriate impact pathways, describing how the mission goals are linked to the inputs provided by a mission (structured along the chain of Inputs-Outputs-Outcomes-Impacts (I-O-O-I)). The development of these pathways through the mission owners and ideally incorporating insights from the systems analysis and stakeholders prevents missions from ending up as a compilation of seemingly related policies and forms the foundation for tracking the progress of a mission and accounting for feedback loops. Thereby the derived impact pathways ensure a shared vision about how to translate mission goals into activities among mission owners and involved stakeholders, and ensures that evaluators have a starting point for their analysis by the description of a sound intervention logic. A key feature of the pathways is thereby the acknowledgement of a weakening control of mission owners over the potential outcomes and impacts (Belcher et al. 2020; Helman et al. 2020). Whereas immediate inputs like policy instruments/activities and their outputs can be shaped by the mission owner (sphere of control), their ability to influence outcomes (sphere of influence) and impacts (sphere of interest) appears to be more limited, as mission activities interact with other elements of the socio-technical system.

The decomposition of complex missions in multiple impact pathways can help to structure the understanding of missions. In order to support this process, we propose a total of eleven stylized pathways that are considered as pivotal in mission realization and that draw on different theory-based strands of research such as transition studies, technological innovation systems (e.g. Ghosh et al. 2021; Wesseling et al. 2020), current work on crafting impact assessment concepts in the context of science and technology (Helman et al. 2020), and empirical insights of the research team of the scientific support action to the German High-Tech Strategy. In contrast to earlier discussions on the economic impact of science (e.g. Salter et al. 2001) and more recent discussions focusing on societal impacts (e.g. Muhonen et al. 2019), we assume that the societal impact of missions in many instances may be stimulated by science, but in others will also be conditional on a wider array of activities. Table 2 provides an overview of the proposed stylized pathways that can serve as a starting point for context-specific impact pathways and their potential relevance for different types of missions.

Table 2: Impact pathways and relation to different mission types (Own elaboration)

	A1	A2	T1	T2
P1: Research to solve problems through targeted research funding	X		X	X
P2: (Basic) Research to generate knowledge for better understanding of the problem	X			X
P3: Collective intelligence/promoting academic exchange to create new knowledge	X	X	X	X
P4: Modification of the research process for better/faster/more solid results	X	X	X	X
P5: Opportunities for new solutions/approaches through positive incentives		X	X	X
P6: Improving framework conditions to increase absorptive capacity		X	X	X
P7: Bringing knowledge & technological approaches to application through targeted support		X	X	X
P8: Creating markets for promising solutions as an impetus for system change			X	X
P9: Exnovation/destabilization of existing regimes to create space/opportunities for new solutions				X
P10: Raising awareness and changing public perceptions (as a prerequisite for change)				X
P11: Change practices, attitudes and behavior to support system changes				X

For example, Pathway P4 targets a modification of the way research is carried out to generate scientific knowledge at a systemic level. Driving motivations can be the aim to increase the quality/robustness of scientific results or better link research activities with societal needs. This may be achieved through the development of new or the adjustment of existing funding schemes, introduction of additional requirements, adjusted peer-review procedures, capacity building, promotion of approaches like citizen science or responsible research and innovation, etc. at the input level. These measures in turn may facilitate first a modified way of doing at the programme level (output) that over time spill-over into a different way of conducting research (outcome), which in turn are a prerequisite to the desired impacts.

4) DEFINITION AND INVENTORY OF INSTRUMENT MIX

Complex interventions such as MOIP require a comprehensive and well-designed mix of instruments, purposefully combining and aligning different instruments with each other. This also includes the deliberate design of new instruments addressing gaps and the realignment/re-orientation of existing policies. At the same time, missions entail the challenge of delineating the instrument mix, thus identifying those instruments that are supposed to contribute to the postulated goals and are under control of the mission owners, and specifying the way they contribute to the developed pathways. The establishment of an inventory of mission instruments, thereby making transparent the key features of

the instruments, supports the strategic orientation of the mission owners and forms the foundation for tracking mission progress along the impact pathways. This might be considered as a top-down approach to identify the instrument mix described by Ossenbrink et al. (2019).

5) TRACKING MISSION PROGRESS ALONG IMPACT PATHWAYS

Mission monitoring should from the beginning be thought of as a part of the implementation process, starting from the mission design towards expected impacts. Making use of the impact pathways, one can derive a monitoring system that allows for an assessment whether the mission is on track or requires adjustment, for example through additional inputs or are re-adjustment of policy measures. In line with previous toolbox elements, the responsibility for this activity is closely associated with the mission owners, but also may involve stakeholders and external evaluators bringing in their expertise and capacity for the identification and collection of the relevant data.

Given the importance of contextual embedding of missions, we refrain from proposing a unified set of indicators as the scope and availability may vary considerably among mission. However, the stylized pathways can provide guidance for the development of appropriate indicators by indicating potentially relevant dimensions that may be worth further consideration

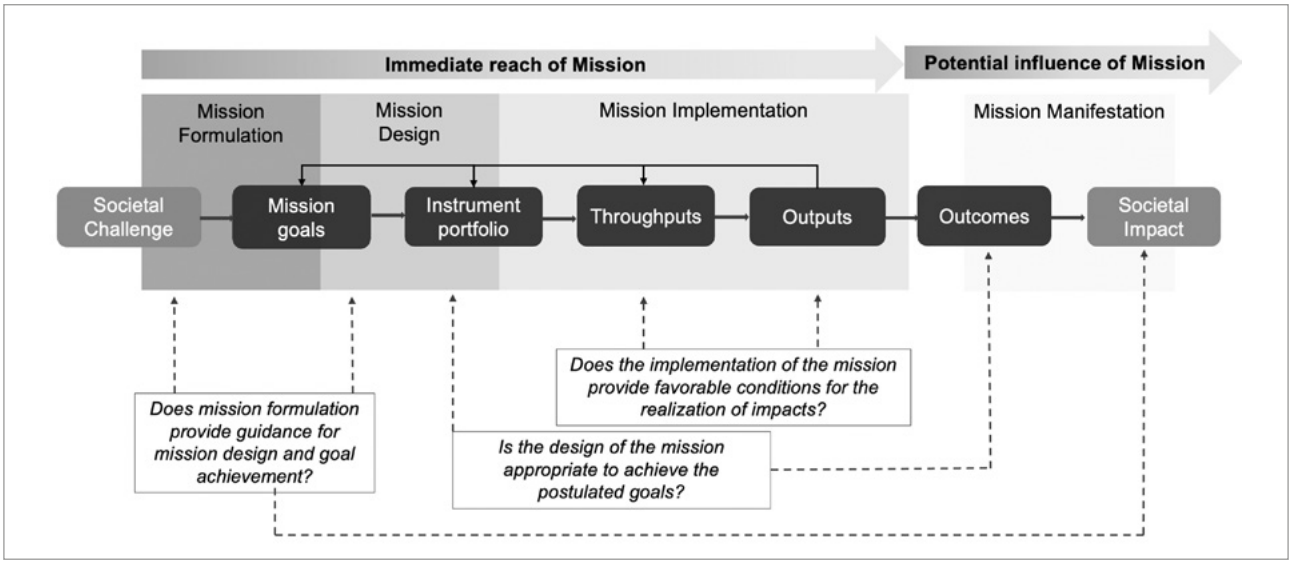
Table 3: Analytical dimensions for indicators (own elaboration)

Input	Output	Outcome	Impact
Incentives/Measures to change established research processes • Awareness raising instruments • Modification of incentives structures (application procedures, requirements etc.) • Dedicated support for key groups or approaches (e.g. citizen science) • Self-declarations and self-commitments	**Modified way of doing research** • Number of projects in supported programmes • Composition of advisory boards/ monitoring bodies • Projects following certain principles/requirements (e.g. RRI) • Funding schemes setting out specific principles/requirements	**Improved results** • Publication, citation, patenting patterns (of underrepresented groups) • Career paths of researchers • Patterns of co-publication, citation, diversity and multidisciplinarity • Uptake in academic debates • Research org. adjusting structures	**Improved knowledge generation** • Robust results through multi-perspectivity • Embedding science into society • More inclusive research

6) ANALYSIS OF TRANSLATION PROCESSES

In contrast to the aforementioned elements of the toolbox that directly interact with the mission owners, the analysis of translation processes explores to what extent favorable or hindering conditions for a materialization of effects were created at different stages. In this regard, it centers on the three guiding questions (see Figure 2 below). Each of these guiding questions encompasses several dimensions, each with a set of more specific analytical questions. Together, the different questions provide for a combination of ex-ante, process- and output-oriented analytical elements. This complementary approach offers a holistic perspective on missions, reaching beyond individual contributions and pathways. Specifically, it is useful for pointing to supporting factors and potential bottlenecks that may hinder mission realization.

Figure 2: Overview of impact assessment concept and guiding questions (own elaboration)



The first analytical question explores whether mission formulation provides sufficient guidance for mission design and possible implementation. This can be considered as an ex-ante element for assessing legitimacy and urgency of missions. Whereas the analysis of goals primarily draws on key strategic documents, the study of mobilization and legitimacy issues may bring together a diverse set of sources, including insights from system analysis, mission typology (see above), expert opinions, participatory observation, and public opinion data. The key analytical dimensions relate to the scope of the mission, definition and operationalization of goals, the relationship between the goals, the legitimacy and ambition of goals, and the mobilization of stakeholders.

Table 4: Analytical questions for mission formulation

MISSION GOAL

Scope of mission
- Does the mission formulate a clear vision/desirable state to be achieved?
- Is the mission explicit in the areas it strives for change/solutions? Does the mission explicitly exclude topics or policy fields? Does the mission contain a justification for its priorities?
- Are mission goals connected to a specific technology? Is the geographical scope of the mission clearly defined?

Definition of goals
- Does the mission have explicitly formulated goals?
- Does the mission include quantitative indicators corresponding to the mission goals? Are mission goals measured on nominal, ordinal, interval or relational scales? Does the mission define a clear baseline/ measurement of the status quo for the intended changes? Does the mission specify data types or sources to be used for measuring goal achievement?
- Do mission goals explicitly define complex constructs linked to goal (e.g. quality of life)? Are mission goals defined in terms of international comparisons or rankings?
- Is a clear time horizon defined for the achievement of mission goals? Does the mission include interim goals or milestones? Do the goals include flexible elements, e.g. if/when context conditions change throughout the mission? Is there a defined process for the adjustment of goals throughout the mission?

Relationship between goals
- Does the mission define more than one goal? Is the prioritization of goals clearly defined?
- Does the mission define if/how one mission goal contributes to other goals? Are postulated goals coherent/non-contradictory or is there a possible tension between goals?

LEGITIMACY, URGENCY, AND PROCESS OF FORMULATION

Legitimacy of goals
- Does the mission refer to a specific societal problem it seeks to address?
- To what extent is there a societal consensus about the importance of the underlying problem? Is there a societal consensus on the urgency of the problem? Do the problems the mission aims to address rank high on the political agenda?

Level of ambition
- Are the mission goals realistic? Are the goals also realistic if context conditions change? Is the realization of mission goals linked to best-case expectations?
- Does the mission aim for transformative change? Do goals go beyond existing trends or push for radical change? Do mission goals appear ambitious compared to similar missions in other countries?

Embedding in political & administrative context
- Is a single mission owner or group of mission owners clearly defined? Can the main mission owner(s) credibly claim capacity/mandate for change (through activities or bringing together relevant actors)? Is the initiator of the mission also responsible for managing the mission?
- Are all relevant political actors and administrative units involved in the mission formulation process? How intense is the collaboration during the mission formulation process? How much attention and support does the mission receive at higher political levels?
- Does the mission refer to existing policies or are there overlapping/duplicating structures at the national level? Does the mission describe how to create synergies based on existing policies? Is it clear what the added value of the mission is, compared to existing policies?
- Does the mission explicitly refer to goals of international strategies? Do the mission goals appear to be in line with international strategies (SDGs, etc.)? Are mission goals aligned with initiatives of supra-national organizations (e.g. EU)?

Legitimacy and stakeholder mobilization
- Are relevant stakeholders (actively) involved in the mission formulation process? Which stakeholders are involved in the process of mission formulation? How are stakeholders identified and selected? Are key stakeholders missing? What are drivers for stakeholders to participate? Are stakeholders incentivized to participate in the mission formulation process?
- Did mission owners reach a mutual understanding of mission goals? Does the involvement of stakeholders include the development of a shared vision? Do stakeholders (formally) commit to the goals formulated?
- Are topical expertise, insights from foresight, or perspectives of stakeholders integrated into the process of mission formulation?
- Is the strategic process of mission formulation designed and equipped with sufficient resources (personnel, financial, temporal)? How does the formulation process deal with possible resistance from key actors/veto players?

The second main question focuses on the ex-ante assessment of the mission design, exploring whether the design of the mission is in line with the expected goals. Essentially, it assesses to what extent the impact pathways and the instrument mix are aligned with the postulated mission goals. At this stage, the analysis mainly draws upon programme documents, expert assessments, insights from the system mapping, participatory observation, and stakeholder perspectives. The analytical questions cover the development of impact pathways, their fit with postulated goals, as well as their consistency and coherence. Further, the questions address the fit between pathways and the instrument mix, as well as the specific character and leverage of policy instruments. Finally, they ask for the process that led to the development of the instrument mix, the coordination of the instruments and the mission's governance structures.

Table 5: Analytical questions for mission design

IMPACT PATHWAYS

Process of pathway development
- Do mission documents (or later provided documents) describe the links between instruments and goals?
- Who lead the process of impact pathway development? To what extent is the development of impact pathways supported by stakeholders or external expertise? What resources are available for the development process?

Fit between pathways & postulated goals
- Are all mission goals addressed by pathways?
- What approach do pathways suggest for achieving the postulated goals? Do goals match with underlying understanding for transformative change?

Consistency of pathways
- Which obstacles need to be overcome to successfully realize the pathways?
- Are pathways appropriate to achieve the desired goals?
- Do pathways aim at second order effects/ cascading effects?
- Coherence of pathways
- Do several impact pathways relate to a shared goal?
- Are there any contradictions/tensions or conflicts arising between pathways?

INSTRUMENT MIX

Fit between pathways & instruments
- Are all impact pathways addressed with instruments/activities? Are pathways highly dependent on one or few dedicated instruments?
- How specific is the alignment of instruments with pathways?

Character of instruments
- What are the main characteristics of the instrument mix applied in the mission (combination of regulation, distribution/incentives, information)?
- Are relevant target groups addressed by the instruments?
- Does the policy instrument mix for individual pathways show gaps or only addresses parts of them?
- Do the mission instruments focus on research output and scientific knowledge production? Do the mission instruments focus on fostering transfer (research to application) and/or adjustment of regulatory frameworks? Do the mission instruments focus on reconfiguring an existing system (e.g. by facilitation of new solutions; building new networks)? Do the mission instruments aim at behavioral change? Do the mission instruments focus on exnovation/regime destabilization/ phase out? Are there compensation mechanisms or incentives for potential losers/actors resisting the anticipated changes?
- Does the policy instrument mix fit the corresponding pathway? Does the instrument mix provide room for experimentation (policy experiments, etc.)? Are there any plans for institutionalizing successful instruments (e.g. pilot projects)?

Leverage of instruments
- What leverage do these instruments possess in the socio-technical system (size, scope, centrality)?
- To what extent does the instrument create synergies with other policies in the field (beyond the mission)? Can the suggested instruments plausibly contribute to a change?

Instrument development & actors' commitment
- Which actors are mobilized to participate in the mission?
- Does the mission mobilize the relevant key stakeholders in the field?
- Which actors are involved in developing the instrument mix? How are instruments identified and selected for the mission? How was the process implemented?
- Who is responsible for instruments of the mission? Were all ministries/public actors active in the field involved in this process? What share of resources is provided by non-public actors?
- Is there a formal commitment of actors to provide resources? How precisely is this defined? Are there incentives for stakeholders to contribute to the mission? To what extent is their contribution formalized?
- Is there a dedicated mission budget?
- Does the commitment include the necessity to adjust/modify existing instruments/activities?
- How is their implementation coordinated between different actors?
- Are the instruments designed specifically for the purpose of the mission or how are existing measures aligned? How were new instruments developed? What resources were available for mission design?

Coordination of instrument mix & governance structure
- What kinds of coordination arrangements are created for the mission? What are their competencies? Who is member of them? How regularly are those planned to convene?
- How is the implementation of instruments coordinated between different actors?
- Are there any pre-defined approaches for mission monitoring, evaluation and learning? How are these to be achieved?

The final group of analytical question deals with the implementation of mission instruments, which represents a central prerequisite for the materialization of impacts in the long run. Examining the implementation of selected instruments and the mission allows to better understand whether favorable conditions exist for the realization of impacts. For this purpose, both the (interim/ex-post) program evaluation of selected (key) instruments as well as the general mission management require an in-depth analysis. Important sources to unravel the implementation of missions are interviews with representatives of administration, stakeholders and experts, as well as participatory observations and document analyses. The key analytical dimensions cover the characteristics of key policies, their effectiveness and potential unintended consequences. Further, they relate to the coordination of the policy mix, as well as the robustness and responsiveness at the implementation stage. Finally, the analytical questions aim at spill-over effects, the quality of monitoring structures and the degree of transparency of mission implementation.

Table 6: Analytical questions for mission implementation

TRANSLATION PROCESSES

Characteristics of key policies
- What are key policy instruments of the mission that are crucial for the success of the mission? To which pathways do they contribute?
- Was the instrument implemented on time? Did the financial volume of the instrument change? Did the policy instruments experience changes in thematic priorities, application regulations etc.?
- Was the programme evaluated?

Effectiveness
- Did the implemented policy instruments have their intended effects?
- Are the instruments implemented in line with the described goals? Was the implementation achieved at reasonable efforts/costs?
- Is there evidence of potential policy-delivery failures?

Unintended consequences
- Did the instrument lead to unintended and undesirable side-effects or secondary effects?
- Did the instruments lead to unintended but desirable side-effects or secondary effects?
- To what extent did learning take place during the implementation process?

MISSION MANAGEMENT

Coordination activities
- What are their competencies and routines (members, main tasks, budget)? Were there additional coordinative bodies created after mission initiation? How regularly do governing/steering bodies of the missions meet?
- Are stakeholders involved in mission governance, e.g. by creation of an advisory board? How are they involved and what are there competencies?

Robustness of implementation
- Were the policy instruments implemented as planned?
- Which policy instruments were terminated or delayed?

Flexibility
- Were policy instruments adjusted? For which reasons?
- Were there any developments/events that would have made a modification of the impact pathways necessary?
- Were instruments able to adapt to exogenous shocks, changing contexts etc.? How fast were instruments adapted? Were adaptive measures
- successful in overcoming obstacles? Is there a regular/scheduled review of the instrument mix and appropriateness of the pathways?

Responsiveness
- How is strategic intelligence (e.g., foresight, evaluations of individual instruments) exchanged within the mission?
- When obstacles or challenges occurred during mission implementation, were the governing/steering bodies able to find and agree on suitable instruments?
- How is the mission progress communicated within the authorities/administration?
- What resources and capacities are available for the coordination of the mission?

Spill-over effects and mobilization
- Does the mission mobilize additional activities/spill-overs for actors that are not part of the mission?
- Does the implementation of the mission contribute to a changing understanding of the underlying problem and its possible solutions for the involved actors and the general public?

Monitoring structures
- Is there a defined process for assessing the progress of the policy instruments of the mission? How regularly is the progress of the instruments assessed?
- Are there defined standards for the reports on instrument progress?
- Is there a clear responsibility to manage the monitoring process? Is there a sufficient budget foreseen for monitoring and evaluation?

Transparency
- Is the progress of the mission/individual regularly discussed at the level of political decision-makers? Is the progress of the instruments part of the mission regularly discussed with stakeholders?
- Are reports on instrument progress regularly communicated to the general public? Is there a unified communication strategy/shared label/website/etc. or does each partner communicate independently?
- How can the outreach of mission activities be assessed?

Feedback & learning
- Does the monitoring feed into the adjustment of instruments?
- Are there processes for collecting experiences/good practices made during mission implementation? Are there structures for institutional knowledge management?
- Is there a process to inform and improve future policies?

DISCUSSION AND CONCLUSION

Our contribution aims not only at proposing a closed framework for impact assessment, but is rather an invitation to all stakeholders interested in empowering the MOIP approach as a means for transformative change. Therefore, we highlight what we consider to be the most important requirements for impact assessment, namely reflecting and acknowledging the need for strong formative, comprehensive and yet integrated perspectives to provide practical guidance to actors involved, and the willingness to follow a theory-based and process-oriented approach to study impacts at the same time. Our flexible and modular toolbox approach pays tribute to these preconditions. It takes the translation processes as a key reference point for analyzing missions, with each translation step accompanied by a set of corresponding questions to guide the assessment. These questions are complemented by additional tools for analysis and assessment. We regard it as important that these tools are developed by or in close cooperation with those implementing the missions and, ideally, with those who are affected by the mission policies. This way, the concept can not only provide for the assessment of missions from the outside, but also support policy-makers, mission owners and stakeholders throughout the mission process.

There are, however, multiple caveats associated with the framework. Embedding the framework into mission implementation and emphasizing the role of formative elements impose high requirements on the involved mission owners. On the one hand, it requires an open administrative culture, willing to actively incorporate stakeholders and seeking close exchange with evaluators that are in charge of the analysis of the translation process. Whereas research has increasingly emphasized the importance of formative evaluation (Magro et al. 2019; Molas-Gallart et al. 2021), there might exist considerable tensions with established working routines and administrative cultures - thus the question whether public actors are willing and institutionally prepared to embrace these principles. On the other hand, the framework is based on the active involvement of different actors and intense reflection processes, entailing significant capacity requirements. The first five toolbox elements aim at making many of the often implicit decision-making processes explicit, thereby supporting the implementation processes. While not providing a blueprint, the tool box elements underline that MOIP does not come at zero costs, but are a highly demanding approach (cf. Lindner et al. 2021).

LITERATURE

Amanatidou, E.; Cunningham, P.; Gök, A.; Garefi, I. (2014): Using Evaluation Research as a Means for Policy Analysis in a 'New' Mission-Oriented Policy Context. In: Minerva, 52 (4), pp. 419–438.

Arnold, E. (2019): Evaluating Complex Innovation and Transition Programmes (CITPs). Manchester.

Arnold, E.; Aström, T.; Glass, C.; Scalzi, M. de (2018): How should we evaluate complex programmes for innovation and socio-technical transitions? Brighton: technopolis group.

Belcher, B. M.; Davel, R.; Claus, R. (2020): A refined method for theory-based evaluation of the societal impacts of research. In: MethodsX, 7, p. 100788.

Borrás, S.; Edler, J. (Eds.) (2014): The Governance of Socio-Technical Systems. Explaining Change. Cheltenham: Edward Elgar Publishing.

Bührer, S.; Reidl, S.; Schmidt, E. K.; Palmen, R.; Striebing, C.; Groo, D. (2019): Evaluation Framework for Promoting Gender Quality in Research and Innovation: How does gender equality influence research and innovation outcomes and what implicants can be derived for suitable evaluation approaches. In: Fteval - Journal for Research and Technology Policy Evaluation, (49), pp. 140–145.

Bührer, S.; Walz, R.; Seus, S.; Astor, M.; Stehnken, T.; Malik, F. (2020): Evaluation der BMBF-Rahmenprogramme Forschung für die Nachhaltigkeit FONA 1 (2005-2009) & Forschung für Nachhaltige Entwicklungen FONA 2 (2010-2014). Abschlussbericht. Studie/Gutachten im Auftrag des Bundesministeriums für Bildung und Forschung (BMBF). Karlsruhe, Berlin: Fraunhofer ISI in Zusammenarbeit mit Prognos.

Cavill, N.; Richardson, D.; Faghy, M.; Bussell, C.; Rutter, H. (2020): Using system mapping to help plan and implement city-wide action to promote physical activity. In: Journal of public health research, 9 (3).

Edler, J.; Berger, M.; Dinges, M.; Gök, A. (2012): The practice of evaluation in innovation policy in Europe.

Edler, J.; Salas Gironés, E. (2020): How do framing and ideas influence the design of missions? A comparative analysis between Germany, the Netherlands, & the United Kingdom. EU-SPRI virtual session on Shaping System Transitions - Insights from practice. 5th of June 2020.

Ghosh, B.; Kivimaa, P.; Ramirez, M.; Schot, J.; Torrens, J. (2021): Transformative outcomes: assessing and reorienting experimentation with transformative innovation policy. In: Science and Public Policy.

Grillitsch, M.; Hansen, T.; Coenen, L.; Miörner, J.; Moodysson, J. (2019): Innovation policy for system-wide transformation: The case of strategic innovation programmes (SIPs) in Sweden. In: Research Policy, 48 (4), pp. 1048–1061.

Griniece, E.; Sorokins, J. (2018): Analysis Report. Responses to the call for feedback on "Mission-Oriented Research and Innovation in the European Union" by Mariana Mazzucato. European Commission - Directorate-General for Research and Innovation.

Hekkert, M. P.; Janssen, M. J.; Wesseling, J. H.; Negro, S. O. (2020): Mission-oriented innovation systems. In: Environmental Innovation and Societal Transitions, 34, pp. 76–79.

Helman, A.; Barberis, M.; Vignetti, S.; Catalano, J.; Griniece, E.; Kroll, H.; Zenker, A.; Martin, C. (2020): Deliverable 5.1. Validated IA Model. Research Infrastructure imPact Assessment paTHwayS.

Hüsing, B.; Kulicke, M.; Wydra, S.; Stahlecker, T.; Aichinger, H.; Meyer, N. (2017): Evaluation der „Nationalen Forschungsstrategie

BioÖkonomie 2030". Wirksamkeit der Initiativen des BMBF - Erfolg der geförderten Vorhaben - Empfehlungen zur strategischen Weiterentwicklung. Abschlussbericht. Karlsruhe.

Janssen, M. J. (2016): What bangs for the bucks? Assessing the design and impact of transformative policy. Center for International Development at Harvard University.

Janssen, M. J.; Torrens, J.; Wesseling, J.; Wanzenböck, I.; Patterson, J. (2020): Position paper. 'Mission-oriented innovation policy observatory'. Utrecht: Copernicus Institute of Sustainable Development, Utrecht University.

Joly, P.-B.; Gaunand, A.; Colinet, L.; Larédo, P.; Lemarié, S.; Matt, M. (2015): ASIRPA: A comprehensive theory-based approach to assessing the societal impacts of a research organization. In: Research Evaluation, 24, pp. 440–453.

Joly, P.-B.; Matt, M. (2017): Towards a new generation of research impact assessment approaches. In: The Journal of Technology Transfer, 1 (4).

Joly, P.-B.; Matt, M.; Robinson, D. K. R. (2019): Research Impact Assessment. From ex post to real-time assessment. Wien: fteval - Platform for Research and Technology Policy Evaluation.

Kalpazidou Schmidt, E.; Bührer, S.; Schraudner, M.; Reidl, S.; Müller, J.; Palmen, R.; Haase, S.; Graversen, E. K.; Holzinger, F.; Striebing, C.; Groó, D.; Klein, S.; Rigler, D.; Høg Utoft, E. (2017a): Conceptual Evaluation Framework for Promoting Gender Equality in Research and Innovation. Toolbox I - A synthesis report. EFFORTI - Deliverable 3.3.

Kalpazidou Schmidt, E.; Cacace, M. (2017b): Addressing gender inequality in science: the multifaceted challenge of assessing impact. In: Research Evaluation, 26 (2), pp. 102–114.

Kroll, H. (2019): How to evaluate innovation strategies with a transformative ambition?: A proposal for a structured, process-based approach. In: Science and Public Policy, 46 (5), pp. 635–647.

Kuittinen, H.; Skov Kristensen, F.; Pelkonen, A.; Lehenkari, J.; Goetheer, A.; van der Zee, F.; Arrilucea, E.; Unger, M.; Türk, A.; Polt, W.; Fisher, R.; Domini, A.; Chicot, J.; Terziev, N.; Vincze, M.; Taranic, I.; Lykogianni, E.; Misojcic, M. (2018): Mission-oriented research and innovation. Assessing the impact of a mission-oriented research and innovation approach: Final report. Luxembourg: European Commission - Directorate-General for Research and Innovation.

Larrue, P. (2021): The design and implementation of mission-oriented innovation policies. A new systemic policy approach to address societal challenges. Paris: OECD.

Lindner, R.; Edler, J.; Hufnagl, M.; Kimpeler, S.; Kroll, H.; Roth, F.; Wittmann, F.; Yorulmaz, M. (2021): Missionsorientierte Innovationspolitik. Von der Ambition zur erfolgreichen Umsetzung. Karlsruhe: Fraunhofer Institut für System- und Innovationsforschung ISI.

Magro, E.; Wilson, J. R. (2019): Policy-mix evaluation: Governance challenges from new place-based innovation policies. In: Research Policy, 48 (10), p. 103612.

Matti, C.; Corvillo, J.; Lalinde, I. (2020): Challenge-led System Mapping. A knowledge management approach.

Mazzucato, M. (2017): Mission-Oriented Innovation Policy. Challenges and opportunities. London: RSA Action and Research Center.

Miedzinski, M.; Allinson, R.; Arnold, E.; Cassingena Harper, J.; Doranova, A.; Giljum, S.; Griniece, E.; Kubeczko, K.; Mahieu, B.; Markandya, A.; Peter, V.; Ploeg, M.; Stasiakowska, A.; van der Veen, G. (2013): Assessing environmental impacts of Research and Innovation Policy. Study for the European Commission, Directorate-General for Research and Innovation. Brussels: technopolis group.

Molas-Gallart, J.; Boni, A.; Giachi, S.; Schot, J. (2021): A formative approach to the evaluation of Transformative Innovation Policies. In: Research Evaluation, Online before print.

Muhonen, R.; Benneworth, P.; Olmos-Peñuela, J. (2019): From productive interactions to impact pathways: Understanding the key dimensions in developing SSH research societal impact. In: Research Evaluation, 14 (1).

Ossenbrink, J.; Finnsson, S.; Bening, C. R.; Hoffmann, V. H. (2019): Delineating policy mixes: Contrasting top-down and bottom-up approaches to the case of energy-storage policy in California. In: Research Policy, 48 (10).

Polt, W.; Weber, M.; Biegelbauer, P.; Unger, M. (2019): Matching type of mission and governance in mission-oriented R&I policy: conceptual improvement and guidance for policy. Eu-SPRI Conference. Rome, 06.06.2020. Available at https://www.researchgate.net/publication/334277744_Matching_type_of_mission_and_governance_in_mission-oriented_RI_policy, accessed 25.09.2019.

Robinson, D. K.; Mazzucato, M. (2019): The evolution of mission-oriented policies: Exploring changing market creating policies in the US and European space sector. In: Research Policy, 48 (4), pp. 936–948.

Salter, A. J.; Martin, B. R. (2001): The economic benefits of publicly funded basic research: a critical review. In: Research Policy, 30 (3), pp. 509–532.

Wanzenböck, 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), pp. 474–489.

Weber, M.; Polt, M. (2014): Assessing mission-orientated R&D programs: combining foresight and evaluation. In: Fteval - Journal for Research and Technology Policy Evaluation, (39), pp. 5–10.

Wesseling, J.; Meijerhof, N. (2020): Development and application of a Mission-oriented Innovation Systems (MIS) approach.

Wittmann, F.; Hufnagl, M.; Lindner, R.; Roth, F.; Edler, J. (2021a): Governing varieties of mission-oriented innovation policies: A new typology. In: Science and Public Policy, 48 (5), pp. 727–738.

Wittmann, F.; Hufnagl, M.; Roth, F.; Yorulmaz, M.; Lindner, R. (2021b): From mission definition to implementation: Conceptualizing mission-oriented policies as a multi-stage translation process. Karlsruhe: Fraunhofer Institut für System- und Innovationsforschung ISI.

Wittmann, F.; Yorulmaz, M.; Hufnagl, M. (2021c): Impact Assessment of Mission-Oriented Policies. Challenges and overview of selected existing approaches. Project deliverable. Karlsruhe: Fraunhofer Institut für System- und Innovationsforschung ISI.

AUTHORS

FLORIAN WITTMANN
Fraunhofer Institute for Systems and Innovation Research
Breslauer Str. 48, 76139 Karlsruhe, Germany
E: florian.wittmann@isi.fraunhofer.de
https://orcid.org/0000-0002-9890-6091

FLORIAN ROTH
Fraunhofer Institute for Systems and Innovation Research
Breslauer Str. 48, 76139 Karlsruhe, Germany
E: florian.roth@isi.fraunhofer.de
https://orcid.org/0000-0002-5287-6448

MIRIAM HUFNAGL
Fraunhofer Institute for Systems and Innovation Research
Breslauer Str. 48, 76139 Karlsruhe, Germany
E: miriam.hufnagl@desy.de
https://orcid.org/0000-0003-0106-6653

Deutsches Elektronen-Synchrotron DESY, Hamburg
Notkestraße 85, 22607 Hamburg, Germany

RALF LINDNER
Fraunhofer Institute for Systems and Innovation Research
Breslauer Str. 48, 76139 Karlsruhe, Germany
E: ralf.lindner@isi.fraunhofer.de
https://orcid.org/0000-0002-6927-8866

MERVE YORULMAZ
Fraunhofer Institute for Systems and Innovation Research
Breslauer Str. 48, 76139 Karlsruhe, Germany
E: merve.yorulmaz@isi.fraunhofer.de
https://orcid.org/0000-0001-7810-4860

KEYWORDS: mission-oriented innovation policy, impact assessment; impact pathways; formative evaluation

DEVELOPING CRITERIA OF SUCCESSFUL PROCESSES IN CO-CREATIVE RESEARCH
A FORMATIVE EVALUATION SCHEME FOR CLIMATE SERVICES

SUSANNE SCHUCK-ZÖLLER, SEBASTIAN BATHIANY, MARKUS DRESSEL, JULIANE EL ZOHBI, ELKE KEUP-THIEL, DIANA RECHID AND MIRKO SUHARI

DOI: 10.22163/fteval.2022.541

ABSTRACT

Climate change and its socio-ecological impacts affect all sectors of society. To tackle the multiple risks of climate change the field of climate services evolved during the last decades. In this scientific field products to be applied in practice are developed in constant interaction between climate service providers and users. To judge the effectiveness of these co-creation endeavours, evaluation is crucial. At present, output and outcome assessments are conducted occasionally in this research field. However, the summative evaluation does not help to adjust the ongoing process of co-creation. Thus, our work focuses on the formative evaluation of co-creative development of science-based climate service products.

As the first step, main characteristics of the product development process were identified empirically. Secondly, we determined the six sub-processes of climate service product development and related process steps. Thirdly, we selected the questions for the formative evaluation relevant to all the sub-processes and process-steps. Then, a literature review delivered the theoretical background for further work and revealed further quality aspects. These aspects from literature were brought together with our results from the empirical work. In the end, we created a new scheme of quality criteria and related assessment questions for the different sub-processes in climate services, based on both, empirical and theoretical work.

As the authors take into account the process of co-production in a real-life case, the criteria and assessment questions proposed are operational and hands-on. The quality aspects refer to the five principles of applicability, theoretical and empirical foundation, professionalism, transparency of processes and the disclosure of preconditions. They are elaborated comprehensively in our study. The resulting formative evaluation scheme is novel in climate service science and practice and useful in improving the co-creation processes in climate services and beyond.

1. INTRODUCTION

Negative impacts of climate change are becoming more apparent. As climate change affects all sectors of society, the cooperation between civil society, political, economic and scientific actors is a key element for the development of climate mitigation and adaptation strategies. For more than fifteen years, the concept of climate services has been developed and regarded as an answer to climate change challenges (Brasseur & Gallardo, 2016). Climate service institutions have been established in many countries.

Climate services are defined as "the transformation of climate-related data — together with other relevant information — into customised products such as projections, forecasts, information, trends, economic analysis, assessments (including technology assessment), counselling on best practices, development and evaluation of solutions and any other service in relation to climate that may be of use for the society at large." (European Commission, 2015, p.10)

Researchers from very different research fields, such as physics, meteorology, biology, agricultural research, social science, economics and others are working together in climate services. They apply the participatory and interactive modes of scientific knowledge production in a highly interdisciplinary research agenda (European Commission, 2015). This agenda comprises approaches of collaboration that includes practitioners. In order to enhance the adaptive capacity of society to climate change, knowledge integration is crucial. Researchers of all relevant fields, partners from practice, and users of climate services need to share their knowledge and learn from each other.

We are aware of multiple and diverse approaches of such interactive modes of scientific knowledge production (for an overview see Newig et al., 2019; Bremer & Meisch, 2018; Brinkmann et al., 2015). Mauser et al. (2013) proposed "co-creation" as an overarching term for the different phases of transdisciplinary research processes in the field of Earth sciences. We follow this proposition and use the terms "co-creation" and "co-creative" research synonymously.

Who might be involved in co-creative projects? On one hand, there are interdisciplinary experts from different research fields, here called "researchers", who are part of a project consortium. They work together in a "scientific team" and form the "scientific party" of the co-creation partnership. The institutions, the researchers belong to, are called climate service "providers". They aim for appropriate information, products and tools to support decision-making in terms of climate change and

impacts. If the researchers need scientific knowledge or data to work with, they might look for "scientific cooperation partners", who are not yet part of the project consortium. They might also involve experts from practice, called "practitioners", to benefit from their experience. They are very often "users" of climate services, who are highly interested in the future products and, therefore, form the "party of practitioners". Beyond the party of practitioners there might be future "external users" as well.

To make co-creation processes effective, their quality and evaluation have been studied for years. In the literature, a broad range of exemplary evaluations assessing whole projects can be found – e.g. in environmental research (Jahn & Keil, 2015), hydrology (Maag et al., 2018) or especially related to weather and climate services (Wall et al., 2017). More in-depth ex post evaluations of single products are still rare, above all in climate services (Körner & Lieberum, 2014; Haße & Kind, 2019). Belcher et al. (2021) only recently have contributed with a quality assessment framework for social innovation impacts in co-creative research.

As evaluation research is increasingly indicating a relation between good co-creation processes and their effectiveness and success (Lux et al., 2020; Maag et al., 2018; Wolf et al., 2013), the adjustment and improvement of these processes become a key issue, above all in climate services, reacting to the urging problem of climate change. Formative evaluation delivers the chance to adapt the processes over the course of an ongoing project and restructure it.

In consideration of the fact that research processes in co-creation are crucial for societal impacts (Maag et al., 2018) the project NorQuATrans (Normativity, Objectivity and Quality of Transdisciplinary Processes) was initiated and implemented[1] in the Helmholtz-Institute HICSS where the scientists from the Climate Service Center Germany (GERICS) and Hamburg University cooperate. The objective of the NorQuATrans is – amongst others – to empirically analyse challenges and limitations of climate services and to examine closely the quality of the co-creation processes. As accompanying research to other projects, the NorQuATrans aims at developing a concept for a formative evaluation scheme of co-creation in climate services as well as a suitable set of quality criteria and methods.

The work described here uses the case study approach by accompanying a co-creative research project in the field of agriculture. The ADAPTER (ADAPT tERrestrial systems) project[2] delivers the innovative simulation-based products to support optimal adaptation to both short-term weather variability and weather extremes, as well as to long-term regional climate change. To do so, ADAPTER involves the practitioners from different areas of agricultural practice. The co-creation processes between ADAPTER scientists and practitioners were used as a showcase to analyse their quality and develop possible evaluation criteria. Hence, this contribution does not present a concrete evaluation activity. It describes instead the development of a methodological approach, leading to a formative evaluation framework aiming at the provision of criteria and related assessment questions for upcoming process evaluations.

To investigate this in more detail, we focus on the joint production of single climate service products. Within the research projects such single product development endeavours are usually just one part of the whole project. Still, our work described here focuses on a consequent close-up view of one single product development process.

The two main questions we aim to answer in this paper are:

- How can climate service providers ensure quality and success of the co-creative processes?
- What are appropriate questions to reflect and evaluate the ongoing processes throughout different project phases?

This is the first study to have developed a formative evaluation scheme dedicated to the field of climate services. Related to the process of climate service product development we apply a method consisting of three steps: a) developing evaluation questions based on newly identified sub-processes of the whole product development endeavour; b) collecting criteria and indicators from a literature overview; c) bringing these deliverables together and adjoining five main principles of co-creative research in climate services.

On this basis, we introduce a process-oriented and formative evaluation scheme encompassing evaluation criteria and assessment questions. The aim is to raise awareness of the importance of formative evaluation amongst researchers or managers in scientific co-creation. It is intended to trigger self-reflection and help the co-creating parties to reflect on the collaboration processes in climate services. Therefore, we adopt and extend existing evaluation schemes for this particular research field.

In the second section, we will present the three-step methodological approach. The results will be presented in the third section, including the process-oriented, formative evaluation scheme. In the fourth section, the results are summarised and discussed in context of the existing literature followed by the limitations of the methodology. Finally, we provide an outlook on open questions and possible further activities in terms of formative evaluation research in the fifth section.

2. METHODOLOGY

The work described here was performed as an accompanying research to the project ADAPTER. Within this project researchers are in dialogue with key agricultural practitioners in Germany. The practitioners' party includes, for example, the farmers' associations on the local level as well as seed grower companies. The collaboration between science and practice within the project has been characterised by intense and regular correspondence by email, by phone and in person. In addition, several workshops took place and at the moment of writing the first product is nearly finished and in the last phase of testing.

THE EMPIRICAL PART

Together with the scientific party in ADAPTER, the NorQuATrans scienific team looked back on the past project phases and identified the steps and sub-steps of their co-creative product development. As a preliminary result, this empirical work firstly showed a workflow containing steps in different grades of details. This workflow – evolved from the ADAPTER project – consisted of a large amount of complexity and details. Therefore, additional experienced climate service researchers were gathered for jointly re-condensing the workflow. As a result, a list of six

1 https://www.hicss-hamburg.de/projects/NorQuATrans/index.php.en
2 www.adapter-projekt.de

sub-processes is summarised. This final list was again revised after the discussion with peer groups (scientists from climate services) and made more coherent.

Then, the scientific parties of ADAPTER and NorQuATrans reflected together how high quality can be ensured during the different process-steps and sub-processes. We adjusted the related evaluation indicators and questions. Particularly in those cases where collaboration had not worked perfectly, we could well identify missing process steps and deduce respective evaluation questions. Intentionally, the six sub-processes and the related questions were developed only from experience.

THE THEORETICAL PART

In parallel, but properly separated, the scientific team of NorQuATrans reviewed the existing literature on formative evaluation. In total, 25 articles from the peer-reviewed journals were identified, which illustrate quality criteria and indicators related to co-creative research. These publications were then analysed for their content and only those which complied with the requirements below were selected:

- focus on the processes – those which focused on output or outcome (OECD, 2002) were dismissed, and
- explicit discussion of co-creative research.

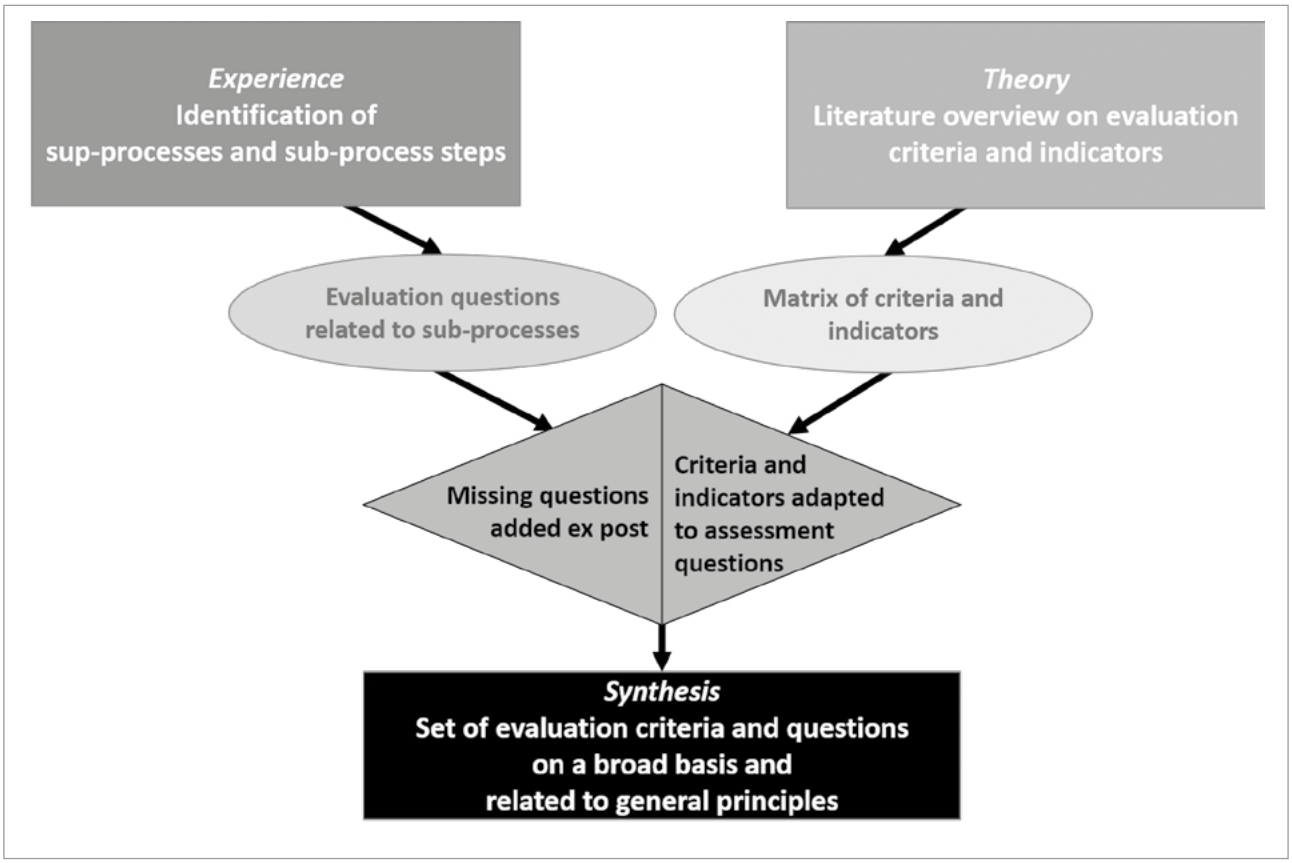
The selected 16 articles presented concrete quality criteria of co-creative research processes – of course with different grades of details – and included different viewpoints. Most of the criteria and indicators were deduced from real cases in practice (e.g. Maag et al., 2018; Bergmann et al., 2005; Wall et al., 2017). A few of the publications analyse huge numbers of projects (e.g. Lux et al., 2020; Newig et al., 2019). No distinction was made during the review whether process criteria and indicators had been identified ex post or during the respective processes. The ADAPTER project team was not involved in this theoretical work.

SYNTHESIS

In the end, the process-related evaluation questions gained by practical experience in climate services and the matrix of quality criteria and indicators from the literature review were compared and synthesized (Fig. 1). Many of the evaluation questions could be directly related to quality criteria and indicators from the theoretical discussion.

If criteria from the literature (here labelled as "theory") were not yet covered by assessment questions from our empirical study ("experience"), suitable questions were added. The criteria and indicators from theory were reformulated if they were not tangible enough to cover our assessment question.

Figure 1: Synthesis of the empirical insights from experiences in the ADAPTER project and theory resulting in a set of evaluation criteria and questions for formative evaluation



The final list of evaluation criteria and questions was sorted and the criteria were aligned with more general categories (here called "principles"), which were also derived from the literature overview.

The two scientific teams a) the ADAPTER co-creative team and b) the NorQuATrans team of accompanying research had different tasks in the course of this development work: Namely the identification of co-creative processes (paragraph a) substantially originated from ADAPTER, the literature review and synthesis work were carried out by the NorQuA-Trans team. A common discussion did not happen until after the synthesis draft was developed and had to be validated.

The development work, presented here, was performed by scientific parties. To pursue the idea of co-creation consequently, practitioners get the chance to contribute and are asked for ex post validation (see section 5).

3. RESULTS

In the following, the results derived from the exercises described above are presented.

SIX SUB-PROCESSES FORM THE CO-CREATION PROCESS IN CLIMATE SERVICES

The empirical work in ADAPTER allowed for the identification of six sub-processes in co-creation (Fig. 2). These sub-processes can be generalised in the direction of a common co-creative process of climate service product development, as they underwent different iteration steps as described above.

RESULTS FROM THEORY AND PRACTICE COMPLEMENT EACH OTHER

The assessment questions gained from the six sub-processes and related process steps were brought together with process evaluation criteria identified through the literature review. This synthesis made clear that the aspects of good quality from theory or practice show many consistencies but complement each other as well. Thus, the collection of criteria resulting from the theoretical review showed gaps in our set of empirical questions. And vice versa, we missed criteria in literature that are seen crucial in practice.

FIVE MAIN PRINCIPLES AND SEVENTEEN QUALITY CRITERIA

To better structure the criteria and related assessment questions, we looked into the literature to find more general terms for categorisation. However, the terms for categories tend to differ. For instance, it is common to see "principles" (Norström et al., 2020; Krause & Schupp, 2019; Belcher et al., 2016) and "dimensions" (Jahn & Keil, 2015), alongside the more normative "recommendations" (Lux et al., 2020; Nagy et al., 2020) or even "elements of success" (Garard et al., 2018) and "areas of improvement" (Jagannathan et al., 2019). We adopted the term "principles", which means here "characteristic principles of co-creation".

Five principles were defined: (1) *common ground*, (2) *transparency*, (3) *professionalism*, (4) *enhancement of applicability* as well as (5) *theoretical and empirical foundation* (table 1). All principles are of equal importance and therefore, the sequence is arbitrary. The second column of table 1 contains the seventeen criteria that we propose. They show what is behind the principles and make them more tangible. Some have already been defined in previous studies. However, we have in some cases changed the wording to make the criterion more suitable for practical use. The four criteria in bold font are newly added by us. As the criteria are overlapping in a few cases, the assignment presented here is not the only sensible possibility.

The corresponding assessment questions will be presented later in the tables 2 to 8. Often one criterion covers different questions, therefore the criteria show up several times.

Figure 2: Six sub-processes of co-creating a climate service product

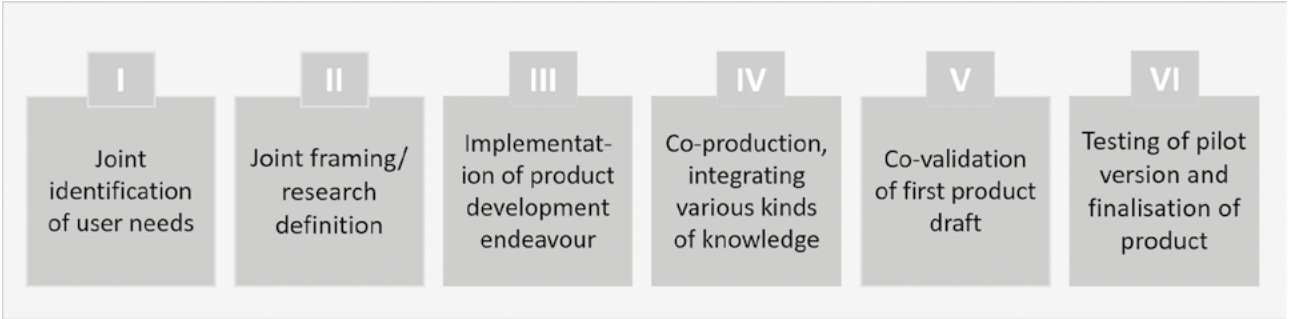


Table 1: Principles and criteria for co-creation processes (by the authors newly added criteria in bold font)

Principles	Criteria	Criterion inspired by	Application in all sub-processes
Common Ground	Equitable opportunities to participate	Belcher et al., 2016; Norström et al., 2020	X
	Trust building	Schuck-Zöller et al., 2018; Norström et al., 2020	
	Joint problem ownership	Schuck-Zöller et al., 2018	
Transparency	Overall development process documentation	Schuck-Zöller et al., 2018	X
	Open exchange on all preconditions		
Professionalism	Clear management and integration concept	Bergmann et al., 2005	X
	Accountability	Krause & Schupp, 2019	X
Enhancement of applicability	Ongoing monitoring and reflection	Bergmann et al., 2005; Belcher et al., 2016	X
	Knowledge integration	Newig et al., 2019; Lux et al., 2020	X
	Benefit of diversity	Maag et al., 2018; Norström et al., 2020	
	Sustainable use		
Theoretical and empirical foundation	Clear research problem definition	Bergmann et al., 2005	
	Scientific soundness	Jahn & Keil, 2015	X
	Use of state-of-the-art knowledge	Belcher et al., 2016; Maag et al., 2018	X
	Scientific peer reviews		
	Coherence	Schuck-Zöller et al. 2018	X
	Impact on science	div.	

Maag et al. (2018) are the first to introduce the term **common ground**, which extends the generally identified need for the category "trust" (Norström et al., 2020; Krause & Schupp, 2019; Schuck-Zöller et al., 2018). Comparing to the term "trust" the notion of *common ground* is more precise to describe communication at eye level, "openness as an attitude" (Garard et al., 2018) and "mutual understanding" (Maag et al., 2018). We understand that this aspect covers the different challenges of communication and collaboration. The related criteria can already be found in Norström et al. (2020), Schuck-Zöller et al. (2018) and Belcher at al. (2016).

Transparency in all collaboration issues is quite generally demanded (i.e. Garard et al., 2018; Schuck-Zöller et al., 2018; Belcher et al., 2016; Jahn & Keil, 2015; Bergmann et al., 2005). This aspect seems a key issue in looking on processes. Thus, we made it a principle in our scheme. *Overall process documentation* is firstly recommended by Schuck-Zöller et al. (2018), the *open exchange on all preconditions* is for the first time taken into account in this paper here. Lux at al. (2019) already pointed out on the clarification of the roles, which is one aspect of the criterion proposed here.

The importance of "good management" is addressed in some contributions (i.e. Wooten et al., 2014; Bergmann et al. 2005). Bergmann et al. (2005) describe in detail how to realise it. Most of the studies agree explicitly that good management is crucial (Schuck-Zöller et al., 2018) or implicitly by noting with other similar criteria (i.e. Wall et al., 2017; Jahn & Keil, 2015; Garard et al., 2018). Still, the importance is not always stressed sufficiently. We decided, that **professionalism** is an appropriate principle to underline the overarching character of this quality aspect and to prevent it from being neglected in practice. Krause & Schupp (2019) point to an aspect which might be underestimated so far: *accountability* of all co-creation participants, above all, of the managing team leads to mutual trust. We see it as a very important aspect of *professionalism*.

The main objective of co-creation endeavours is without any doubt the applicability of research results. While this can be reviewed best ex post, we demand the **enhancement of applicability** for the co-creation process. Very common here is the demand for *ongoing monitoring and reflection* (i.e. Bergmann et al., 2005; Belcher et al., 2016) and *knowledge integration* (i.e. Newig et al., 2019; Lux et al., 2020), whereas the *benefit of diversity* is not highlighted very often (except for Maag et al., 2018; Norström et al., 2020). *Sustainable use* is made a criterion of climate service product development here for the first time. We want to stress how essential it is to provide for long-lasting use of products.

The **theoretical and empirical foundation** (Belcher et al., 2016) alludes to the soundness of both, research as well as experiential knowledge and all integration activities. It is strongly related to *professionalism* in research and the facilitation of *knowledge integration*. We follow several papers with the aligned criteria (Jahn & Keil, 2015; Belcher et al., 2016; Maag et al., 2018; Schuck-Zöller et al., 2018). The special demand for *scientific peer reviews* is a special concern of ours, based on the experience that thorough discussions with scientific peers and critical reviews of the ongoing product development often get out of sight. *Impact on science* is adopted from traditional evaluation of basic research. We apply it to co-creative endeavours as well.

Which gaps did we identify by allocating the assessment questions from experience with process evaluation criteria from the literature review? We missed in literature, for instance, informative criteria related to *theoretical and empirical foundation*. This principle was rarely made an issus in terms of co-creation processes quality. Furthermore, the criterion *open exchange on all preconditions* to the upcoming processes was not addressed in this general meaning. In the end, four criteria seemed to be missing and essential enough to be added by us (in bold font, see table 1).

FINDINGS CONDENSED IN A FORMATIVE EVALUATION SCHEME

The synthesis described above leads to a scheme of evaluation criteria and questions, aligned to the different principles of co-creation covering the process from the first idea of a climate service product to the finalised version. The collection of questions is broad and the questions are general enough to be usable for different kinds of products.

How to use the scheme? Our aim is to provide a set of reasonable questions to researchers and co-creation facilitators – regardless whether these two tasks have to be taken over by one person or not – who want to monitor and evaluate a co-creation process. The scheme can either be used by the co-creating researchers or the facilitators of the co-creation endeavour (self-evaluation), by colleagues from their institution, who are not involved in the product development (in-house evaluation), or by external evaluation specialists (external evaluation) who might look neutrally on the development process.

The type of questions and their addressees differ: Some are to be answered by the co-creation facilitating team themselves. Others allow for learning about the perception of either all participants or just the practitioners or researchers. The questions are recommended to be operationalised by a five-point Likert-scale, or as simple as a binary question with Yes/No-answer. The possibility to choose between different gradations can increase the motivation to answer the questions and allows for comparing the perception of different participating parties. If there are time constraints, we recommend not to use open questions, as they require more resources for answering and interpreting. However, open questions reveal more in-depth information. In general, we decided to design the set of questions incoherently in form (i.e. open or closed): The change of different survey techniques may give a more nuanced picture and make the participation more interesting.

In general, the sequence of the questions in the tables follow the course of the co-creation process.

Some of the questions are quite similar to each other on purpose. Different principles and criteria mirror different aspects of quality and might suit for triangulation. The question *Have all practitioners been included in the previous step of reflection?*, for example, focuses on the possibility to participate, whereas *How many practitioners have taken part in the previous step of reflection?* concentrates on the readiness of practitioners to appreciate this possibility. But in general, some of the alignments are propositions and not compelling. The challenging of these alignments could be a first step to the self-reflection we aim to trigger.

The assessment questions in tables 2 to 8 define the underlying indicators. These indicators are presented in bold font. As we want to present a scheme sufficiently general for different kinds of products, some questions might not fit the objective of that very special product development endeavour. We invite the project leaders to select those questions that apply to their development process and state reasons why others do not.

Some of the evaluation questions came up in every sub-process in similar shape, due to the demand for ongoing reflection. To simplify the scheme, we extracted these evaluation criteria and the related 24 evaluation questions and generalised them. We propose to use them in every sub-process (table 2).

Table 2: SUB-PROCESSES I to VI **– Common criteria for formative evaluation**

Principle	Criterion	Question to reveal results for underlying indicators
Common Ground	Equitable opportunities to participate	*Have all participants perceived **balanced opportunities** to participate?*
		*Has a **balanced influence** between all project partners (from science and practice) been assured in this sub-process?*
		*Have **all practitioners been involved in** the previous step of reflection?*
Enhancement of applicability	Knowledge integration	***How many practitioners participated** in the previous step of reflection?*
Theoretical and empirical foundation	Use of state-of-the-art knowledge	*Have the a) **recent data base** and b) **recent analyse methods been used** in the previous research step?*
	Scientific soundness	*Have the tasks in the respective sub-process been executed in a **scientifically sound** manner?*
	Coherence	*Have **contradicting viewpoints** of single project partners been a) **handled constructively** and b) **made coherent**?*
Professionalism	Accountability	*Have the **management methods** in the respective subprocess been **applied appropriately**?*

Transparency	Overall development process documentation	*Has the respective **sub-process** been* *a) **transparent** to all participants and* *b) **properly** documented?*
Enhancement of applicability	Ongoing monitoring and reflection	*Has the **format for reflection** during the upcoming step been chosen **appropriately**?*
	Knowledge integration	*To which extent have in the previous reflection step been **considered*** *a) the original **contributions of knowledge** (e.g. local, experiential) **from practitioners** and* *b) the **feedback of practitioners**?*
		*Have **consequences been taken from the feedback** gained in the former reflection step?*
		*Have all **participating parties been satisfied** with the former reflection step concerning* *a) format,* *b) method and* *c) result?*
		*Has the **application of integration formats and methods** been **successful**?*
Professionalism	Clear management and integration concept	*Have the process steps of the past sub-process been **executed in line with the different plans**, i.e.* *a) the time schedule in detail (mile stones),* *b) the integration concept and* *c) the documentation concept?*
		*Do parts of the **plans need adaptation**?*

These questions should be taken into account in each of the six sub-processes discussed here. Listing different aspects of the questions (e.g. a, b, c, etc.) might help to realise different facets and/or illustrate the criterion.

All other criteria and evaluation questions are presented in the context of the different sub-processes (fig. 2) and follow the workflow within. Sub-process I starts with the idea of a climate service product be it expressed by practitioners, researchers or funding institutions. The most important steps are related to the identification and recruitment of key experts from practice (practitioners) and the specification of their needs. Therefore, the questions mainly focus on these aspects (table 3).

Table 3: SUB-PROCESS I – Joint identification of user needs – **Criteria for formative evaluation**

Principle	Criterion	Question to reveal results for underlying indicators
Enhancement of applicability	Benefit of diversity	*Has the **selection of practitioners** been conducted in a **systematic way** concerning the project content and goals of co-creation?*
	Ongoing monitoring and reflection	*Has the **selection process been reflected** within the project consortium?*
	Benefit of diversity	*Are the **targeted user groups appropriately represented** by the selected practice partners?*
		*Has the **analysis of user needs** been executed **methodologically sound**, i.e.* *a) open-ended,* *b) supported by balanced and appropriate communication and/or* *c) providing balanced opportunities to utter needs?*
	Ongoing monitoring and reflection	*Has the result of the **analysis of user needs been shared** with the practitioners and commonly reflected?*
Common Ground	Trust building	*How far have **trust and motivation been established** during the contact phase with the practitioners?*
		***How many of the desired practitioners** could successfully **be recruited**?*

The eligible disclosure of *all preconditions* for the co-creation endeavour and their *open exchange* follows in sub-process II (table 4). Several questions are to illuminate the different aspects of this newly introduced criterion. Assessing *scientific soundness* in sub-process II leads to a further reflection on the feasibility and methodological limits. To finalise the sub-process, we propose to look back on the *knowledge integration* in general: *How well does the scientific research question cover the needs from practice?*

Table 4: SUB-PROCESS II – Joint framing/research definition – **Criteria for formative evaluation**

Principle	Criterion	Question to reveal results for underlying indicators
Transparency	Open exchange on all preconditions	*To what extent heave **general preconditions**, such as* *a) mutual expectations,* *b) potential benefits and* *c) potential risks* *been **shared between all participants**?*
		*Have all **formal or external preconditions** for a co-creation endeavour been **shared between all participants**, i.e.* *a) readiness for open communication versus restriction as for strategic or competition issues,* *b) timely resources on all parties' sides and/or* *c) financial resources and conditions?*
Professionalism	Clear management and integration concept	Have ***formal or external conditions been taken into account*** for the management and integration concept?
Theoretical and empirical foundation	Use of state-of-the-art knowledge	*Has the **scientific feasibility** been proven, i.e.* *a) are appropriate climate (model)data or model ensembles available?* *b) have scientific state-of-the-art methods already been developed?* *c) will scientific state-of-the-art methods be able to be applied?* *d) have scientific cooperation partners – if needed – been found?* *e) could identified gaps be filled in?*
Transparency	Open exchange on all preconditions	*Has the **proof of scientific feasibility** as well as the potentials and limitations of research methods (i.e. bandwidths in climate simulations) been **shared*** *a) within the scientific team,* *b) with project partners and* *c) with practitioners?*
Common Ground	Trust building	Have ***all practitioners been included*** in the discussion on ***scientific methodology***?
Enhancement of applicability	Ongoing monitoring and reflection	*To which extent have **practitioners** been able to **accept limits of research** methods or other external conditions?*
Theoretical and empirical foundation	Clear research problem definition	*Have* *a) temporal and* *spatial **scales of the scientific answers** aimed for been clearly **defined**?*
	Scientific peer reviews	*Have common **discussions on the formulation of the research question with scientific peers**, i.e. colleagues taken place?*
Enhancement of applicability	Ongoing monitoring and reflection	*Has the **research question** been* *a) thoroughly **discussed with practitioners** and* ***a) formulated jointly**?*
Common Ground	Joint problem ownership	*To which extent is the research question identified with by* *a) practitioners and* *b) scientists?*
Theoretical and empirical foundation	Knowledge integration	*How well does the **scientific research question cover the needs** from practice?*

The most distinctive sub-process is the third one (table 5). The facilitators of the co-creation endeavour have to set up fair and realistic concepts for any sort of management and knowledge integration and, thus, determine the character of the further co-creation. The researchers have to contribute with detailed descriptions of their tasks and the identification of related milestones. Many of the questions and criteria refer to the principle of *professionalism*. Key for this rather challenging sub-process is to include practitioners and all those participating in the decision-making process that will shape the upcoming co-creation.

Table 5: SUB-PROCESS III – Implementation of product development endeavour – **Criteria for formative evaluation**

Principle	Criterion	Question to reveal results for underlying indicators
Professionalism	Clear management and integration concept	*Has a **knowledge integration** concept been established describing* *a) internal communication,* *a) regular reflection after every sub-process,* *b) how the integration of different kinds of knowledge can be supported and* *c) how different methods of co-creative research can be reflected?*
		*Has a **management plan** been set up containing* *a) a time schedule in detail (mile stones) and* *b) evaluation criteria?*
		*Does the **timeframe** of the project **meet all project partners**'* *a) needs,* *b) constraints and* *c) goals?*
		*To which degree does the planning of the product development enable to **react to the results of the different iteration** processes?*
Transparency	Overall development process documentation	*Has a **concept for the documentation** of process steps been established?*
Enhancement of applicability	Ongoing monitoring and reflection	*Have the **scientific team and the practitioners agreed upon** the different **concepts**, i.e.* *a) the time schedule in detail (mile stones),* *b) the integration concept,* *c) the documentation concept,* *d) an external communication concept of outputs and/or* *e) the evaluation criteria?*
Professionalism	Clear management and integration concept	*How **realistic** is the **product development schedule** in general, i.e. including phases of internal communication, reflection and iteration?*
		*Have the different steps of the **scientific process** been **planned thoroughly**, i.e. related to external preparative work?*
Common Ground	Equitable opportunities to participate	*To which extent are the* *a) researchers and* *b) practitioners* *s**atisfied with the level of engagement**?*

Sub-process IV (table 6) is the most complex one and will take the most time. It entails the process of co-producing research results and often takes several months to even years. Though *knowledge integration* is an overarching task of the whole product development process, it is a key aspect here to assure the *applicability* of the climate service product. Therefore, we explicitly stress its importance. Sub-process IV leads to a first draft of the product.

Table 6: SUB-PROCESS IV – Co-production, integrating various kinds of knowledge – **Criteria for formative evaluation**

Principle	Criterion	Question to reveal results for underlying indicators
Theoretical and empirical foundation	Coherence	*To which extent has it been possible to combine **scientific excellence with the aim of solving real-world problems**?*
	Scientific peer reviews	*How far have the different product development steps and their results been **discussed with peers from science**, i.e. by presentations on scientific meetings/conferences?*
	Coherence	*To which extent are the a) researchers and b) practitioners **satisfied with the joint research** process?*
Enhancement of applicability	Ongoing monitoring and reflection	*Has the the **first draft of the product been reflected** jointly with the researchers and practitioners?*
	Knowledge integration	*To which extent has the development team succeeded in **meeting the problems and objections resulting from the common reflection** of the first draft of the product?*

This first product draft has to undergo a thorough testing by the target group (sub-process V, table 7). The conditions for this validation steps have to be designed close to real ones. Consideration should be given to expanding the group of test users to include external potential users. A validation by peers from science is recommended as well. After the different revisions, a pilot version of the product is created.

Sub-process VI (testing of pilot version and finalisation of product) shows similar steps (table 8) like sub-process V: After tests of the pilot version by different user groups, the pilot is revised and brought to application. The first application phase delivers the chance of last revisions before the product will be finalised. In this phase, it is crucial to ensure sustainability by providing an easy-to-use manual. A long-term support might further enhance the chance of *sustainable use.*

Table 7: SUB-PROCESS V – Co-validation of first product draft – **Criteria for formative evaluation**

Principle	Criterion	Question to reveal results for underlying indicators
Enhancement of applicability	Ongoing monitoring and reflection	*Has the **format for the different steps of co-validation been chosen appropriately**?*
		*Is the format for the different steps of **co-validation of the first product draft close to reality**?*
	Knowledge integration	*Is the **first product draft easy to use**?*
	Ongoing monitoring and reflection	*Has **the first product draft** been a) **tested**, b) **revised** and c) **finalised** **with practitioners**?*
Theoretical and empirical foundation	Scientific soundness	*Has the **first product draft** been **validated by peers** from science and **revised** accordingly?*
Enhancement of applicability	Knowledge integration	*Has the project team succeeded in meeting the **problems and objections** resulting from **the co-validation of the first product draft**?*

Table 8: SUB-PROCESS VI – Testing of pilot version and finalisation of product – **Criteria for formative evaluation**

Principle	Criterion	Question to reveal results for underlying indicators
Enhancement of applicability	Ongoing monitoring and reflection	*Is the format for the different steps of **co-validation of the pilot version close to reality**?*
	Benefit of diversity	*Has an appropriate point at which **external target audiences** enter been made an issue?*
	Sustainable use	*Have the needs a) for a **manual and** b) for a **long-term support concept** been **considered**?*
		*Has a l**ong-term support concept** – if needed – been **assured**?*
		*Is there **staff provided to do a long-term support**?*
	Ongoing monitoring and reflection	*Has the **manual** – if needed – been a) **tested**, b) **revised and** c) **finalised**?*
		*Has the **service concept** – if needed – been a) **proven**, b) **revised and** c) **finalised**?*
Professionalism	Accountability	*Is the **pilot version easy to use**?*
Theoretical and empirical foundation	Scientific soundness	*Is the **group to test the pilot version big** and **diverse** enough?*
		*Has the **pilot version** been **tested within a scientific peer group**?*
	Impact on science	*To what extent have the **findings** of the product development endeavour and the **research results contributed** to **the scientific community**?*
Enhancement of applicability	Ongoing monitoring and reflection	*Has the **pilot version** been a) **tested**, b) **revised and** c) **finalised** **with practitioners**?*
	Knowledge integration	*Has the project team succeeded **in meeting the problems and objections** resulting from **the co-validation of the pilot version**?*

It is obvious which criteria are the most important ones: *Ongoing monitoring and reflection* and *knowledge integration* are named most frequently in our tables. They are to be assessed by questions applicable in the same shape over all sub-processes but, in addition, scrutinised by more special questions and indicators during each single sub-process.

4. SUMMARY AND REFLECTIONS

The projects NorQuATrans and ADAPTER identify six sub-processes of climate service product development. Five principles of co-creation as well as seventeen criteria allow for assessing the quality of co-creative development processes. By covering the quality criteria by assessment questions and indicators, easy-to-use application is provided. A framework is presented for climate service researchers, managers and other participating parties to thoroughly reflect on.

We show that all sub-processes of the product development workflow can be addressed by a specific set of evaluation questions and underlying indicators.

We started with an experience-based analysis and then – like Maag et al. (2018), Wall et al. (2017) and Belcher et al. (2016) did – widened the perspective by a literature review that delivered a more general point of view. The synthesis of both, empirical analysis and theoretical background leads to criteria and indicators, which target on very concrete product development processes and should be applicable in practice.

The case study approach is, as well, used by many of the forerunners, like Bergmann et al. (2005), Wall et al. (2017), Maag et al. (2018). We proceeded similarly to Maag et al. (2018), but identified only six sub-processes ending with the finished product, i.e. excluding implementation and dissemination activities. Like Jahn & Keil (2015), we revealed rather concrete aspects. However, in contrast to most of the existing literature, we only investigated one specific development process in detail.

We experienced that the merging of the empirical development work with the discussions in literature generated particularly interesting results by identifying gaps in our set of empirical questions and vice versa. For example, the quality of iteration and reflection steps of both co-creation parties did not seem to be satisfyingly defined so far. Recommendations were missing, how to proceed in details. Thus, we integrated numerous related assessment questions in every single sub-process to stress the need for repeated common reflection.

Our study aims at the development of a methodological approach and leads to a formative evaluation framework. This is the first time that such a scheme is developed for the dedicated use in the field of climate services.

Especially the principle *of theoretical and empirical foundation* is well elaborated by six different criteria – two of them firstly introduced in this context. These six criteria are not only about using state-of-the-art methodology, but about sharing experiences with the scientific community and inspiring other co-creative projects. The fact, that scientific quality is rather rarely made an issue in literature, might originate from the focus on the co-creation dialogue process phases in previous publications.

What we also introduce for the first time, is the highlighting of different kinds of *preconditions* and the trial to grasp them at the beginning of the development process. Furthermore, for the last sub-process the idea of *sustainable use* is emphasized and backed by assessment questions.

The evaluation framework aims at enhancing the readiness for formative evaluation in co-creation processes. This might be an external, an internal or a self-evaluation. Therefore, we do not see a problem or role conflict in having involved the co-creative scientists in this framework development because co-creative researchers have to cover both roles in self-evaluation activities, as well.

For time constraints, the point of view of practitioners could not yet be directly incorporated into the work described. Still, the co-creative scientists contributed with their experience from practitioner dialogues. Thus, the practitioners' view was represented indirectly. A direct reaction by practice parties to the criteria defined is pending.

One could argue that a single project in the focus impedes generalisation. We tried to meet this challenge by repeatedly discussing the development stage with peers from climate services and comparing the experiences from ADAPTER with other product development processes. Therefore, we are convinced that the six sub-processes allow for generalisation in the direction of a common co-creative process of climate service product development. However, future projects of climate services are to validate the workflow and adjust as well as generalise it. A further step in the direction of generalisation will be the validation of criteria by surveys (see section 5). As the literature review covers all fields of co-creative research, it does not only deliver a theoretical confirmation of the empirical questions and underlying indicators, but also enhances their potential to be transferred to other transdisciplinary research areas.

As we present a framework sufficiently general for different kinds of products, some questions might not fit the objective of every special product development endeavour. We invite the project leaders to select those questions that apply to their development process and perhaps add new criteria and assessment questions. Furthermore, before a new development endeavour is initiated, the set of quality criteria should be discussed. If criteria are to be left out, the reasons should be stated and new criteria could be argued for. Thus, the set of questions is adjustable to the objective of the specific product development endeavour.

5. OUTLOOK

To integrate practitioners' view, we are going to have key criteria of this scheme validated by experts from practice. For this purpose, an empirical study is being prepared in NorQuATrans. Hence, the criteria presented here will be further reviewed by experts from practice beyond the ADAPTER project and the agricultural sector. Thus, we will gain more general information on practitioners' understanding of criteria and their priorities. An additional survey with scientists of different fields of co-creation would further enhance the potential for generalisation. Still missing is an application test of the whole evaluation scheme. This has to be performed, once the scheme will be further validated.

Another still widely open field is the issue of "co-evaluation" (Lux et al., 2020) of co-creation processes and beyond. In general, this aspect has not yet been sufficiently studied – at least in climate services, except for Restrepo et al. (2020). In applying this framework, an open discussion on the evaluation concepts, results and possible re-adjustments should be performed by all participants of the co-creation endeavour from the very beginning. Thus, the scheme could ensure *transparency* for all involved actors and throughout all co-creation phases. A consequent disclosure of all perspectives in assessing quality could make a difference for future work in this respect.

Our framework can be expanded into guidelines for formative evaluation in the future. A comprehensive manual on formative evaluation in climate services is to be generated and presented to research organisations and communities of transdisciplinarity and climate research. Finally, the results can benefit other co-creative research fields. We assume that the scheme for formative evaluation presented here, as well as the resulting guidelines, might be transferable – at least in parts – to other fields of co-creative research.

ACKNOWLEGEMENTS

This work was conducted and financed within the framework of the Helmholtz Institute for Climate Service Science (HICSS), a coperation between Climate Service Center Germany (GERICS) and Universität Hamburg, Germany.

This work was supported by the project ADAPTER, which is funded by the Helmholtz Association under agreement number WT-0104.

We would like to thank Laura Schmidt for a comprehensive internal review and Jo-Ting Huang-Lachmann for her very thorough English language check.

REFERENCES

(Publications included in the literature overview are marked by: L)

Belcher, B., Claus, R., Davel, R., Jones, S.M. (2021). Evaluating and improving the contributions of university research to social innovation. Social Enterprise Journal 1750-8614

Belcher, B., Rasmussen K.E., Kemshaw, R.M., Zornes, D. (2016). Defining and assessing research quality in a transdisciplinary context. Research Evaluation 25(1) – L

Bergmann, M., Brohmann, B., Hofmann, E., Loibl, M.C., Rehaag, R., Schramm, E., Voß, J.P. (2005). Quality Criteria of Transdisciplinary Research. A Guide for the Formative Evaluation of Research Projects. ISOE-Studientexte 13. Frankfurt am Main – L

Brasseur, G.P., Gallardo, L. (2016). Climate Services: Lessons Learned and Future Prospects. Earth's Future 4(3)

Bremer, S., Meisch, S. (2017). Co-production in climate change research – reviewing different perspectives. WIREs Climate Change 8 (6)

Brinkmann, C., Bergmann, M., Huang-Lachmann, J.T., Rödder, S., Schuck-Zöller, S. (2015). Zur Integration von Wissenschaft und Praxis als Forschungsmodus. Climate Service Center Germany. Report 23, Hamburg

European Commission (2015). A European Research and Innovation Roadmap for Climate Services. https://ec.europa.eu/programmes/horizon2020/en/news/european-research-and-innovation-roadmap-climate-services (31.10.2021)

Garard, J., Koch, L., Kowarsch, M. (2018). Elements of success in multi-stakeholder deliberation platforms. Palgrave communications 4:129 – L

Haße, C., Kind, C. (2019). Updating an existing online adaptation support tool: insights from an evaluation. Climatic Change 153

Jahn, T., Keil, F. (2015). An actor-specific guideline for quality assurance in transdisciplinary research. Futures 65 – L

Jagannathan, K., Arnott, J.C, Wyborn, C., Klenk, N., Mach, K.J., Moss, R.H., Sjostrom, K.D. (2019). Great expectations? Reconciling the aspiration, outcome, and possibility of co-production. Current Opinion in Environmental Sustainability 42 – L

Körner, C., Lieberum, A. (2014). Instrumente der Anpassungskommunikation in nordwest2050. Evaluation der Online-Medien. Beese, K., Katz, C., Körner, C., Molitor, H. (eds.) Anpassung an regionale Klimafolgen kommunizieren. Konzepte, Herausforderungen und Perspektiven. München, Oekom

Krause, G., Schupp, M. (2019). Evaluating knowledge transfer at the interface between science and society. GAIA 28/3 – L

Lux, A., Schäfer, M., Bergmann, M., Jahn, T., Marg, O., Nagy, E., Ransiek, A.C., Theiler, L. (2020). Societal effects of transdisciplinary sustainability research. How can they be strengthened during the research process? Environmental Science & Policy 101 – L

Maag, S., Alexander, T.J., Kase, R., Hoffmann, S. (2018). Indicators for measuring the contributions of individual knowledge brokers. Environmental Science & Policy 89 – L

Mauser, W., Klepper, G., Rice, M., Schmalzbauer, B.S., Hackmann, H., Leemans, R., Moore, H. (2013). Transdisciplinary global change research: the co-creation of knowledge for sustainability. Current Opinion in Environmental Sustainability (5)

Nagy, E., Ransieka, A., Schäfer, M., Lux, A., Bergmann, M., Jahn, T., Marg, O., Theiler, L. (2020). Transfer as a reciprocal process: How to foster receptivity to results of transdisciplinary research. Environmental Science & Policy 104 – L

Newig, J., Jahn, S., Lang, D.J., Kahle, J., Bergmann, M. (2019). Linking modes of research to their scientific and societal outcomes. Evidence from 81 sustainability-oriented research papers. Environmental Science & Policy 101 – L

Norström, A.V., Cvitanovic, C., Löf, M.F., West, S., Wyborn, C., Balvanera, P., Bednarek, A.T., Bennett, E.M., Biggs, R., de Bremond, S., Campbell, B.M., Canadell, J.G., Carpenter, S.R., Folke, C., Fulton, E.A., Gaffney, O., Gelcich, S., Jouffray, J.B., Leach, M., Le Tissier, M., Martín-López, B., Louder, E., Loutre, M.V., Meadow, A.M., Nagendra, H., Payne, C., Peterson, G.C., Reyers, B., Scholes, R., Speranza, C.I., Spierenburg, M., Stafford-Smith, M., Tengö, M., van der Hel, S., van Putten, I., Österblom, H. (2020). Principles for knowledge co-production in sustainability research. Nature Sustainability 3 – L

OECD (2002). Glossary of key Terms in Evaluation and results based management. http://www.oecd.org/development/peer-reviews/2754804.pdf (31.10.2021)

Pohl, C., Truffer, B., Hirsch Hadorn, G. (2017). Addressing Wicked Problems through Transdisciplinary Research. Robert Frodeman (ed). The Oxford Handbook of Interdisciplinarity (2nd). Oxford University Press

Restrepo, M.J., Lelea, M.A., Kaufmann, B.A. (2020). Assessing the quality of collaboration in transdisciplinary sustainability research. Farmers' enthusiasm to work together for the reduction of postharvest dairy losses in Kenya. Environmental Science & Policy 105

Schuck-Zöller, S., Brinkmann, C., Rödder, S. (2018). Integrating Research and Practice in Emerging Climate Services: Lessons from Other Transdisciplinary Dialogues. Serrao-Neumann, S., Coudrain, A., Coulter, L. (eds). Communicating Climate Change Information for Decision-Making. Springer International – L

Vaughan, C., Dessai S. (2014). Climate services for society: origins, institutional arrangements, and design elements for an evaluation framework. Wiley Interdisciplinary Reviews. Climate Change 5 – L

Wall, T.U., Meadow, A.M., Horganic, A. (2017). Developing evaluation indicators to improve the process of coproducing usable climate science. Weather, Climate, and Society 9 – L

Wolf, B., Lindenthal, T., Szerenscits, M., Holbrook, J.B., Heß, J. (2013). Evaluating Research beyond Scientific Impact. GAIA 22(2) – L

Wooten, K.C., Rose, R.M., Ostir, G.V., Calhoun, W.J., Ameredes, B.T., Brasier, A.R (2014). Assessing and Evaluating Multidisciplinary Translational Teams. A Mixed Methods Approach. Evaluation & the Health Professions 37(1) – L

AUTHORS

SUSANNE SCHUCK-ZÖLLER
Climate Service Center Germany (GERICS)/Helmholtz-Zentrum Hereon
Fischertwiete 1, 20095 Hamburg, Germany
E: susanne.schuck@hereon.de
ORCID ID: 0000-0003-0992-8334

SEBASTIAN BATHIANY
Climate Service Center Germany (GERICS)/Helmholtz-Zentrum Hereon
Fischertwiete 1, 20095 Hamburg, Germany
E: sebastian.bathiany@hereon.de

MARKUS DRESSEL
Climate Service Center Germany (GERICS)/Helmholtz-Zentrum Hereon
Fischertwiete 1, 20095 Hamburg, Germany
E: markus.dressel@hereon.de

JULIANE EL ZOHBI
Climate Service Center Germany (GERICS)/Helmholtz-Zentrum Hereon
Fischertwiete 1, 20095 Hamburg, Germany
E: juliane.el_zohbi@hereon.de
ORCID ID: 0000-0003-0329-0552

ELKE KEUP-THIEL
Climate Service Center Germany (GERICS)/Helmholtz-Zentrum Hereon
Fischertwiete 1, 20095 Hamburg, Germany
E: elke.keup-thiel@hereon.de

DIANA RECHID
Climate Service Center Germany (GERICS)/Helmholtz-Zentrum Hereon
Fischertwiete 1, 20095 Hamburg, Germany
E: diana.rechid@hereon.de
ORCID ID: 0000-0002-6035-2935

MIRKO SUHARI
Leuphana University
21335 Lüneburg, Germany
E: mirko.suhari@leuphana.de

KEYWORDS: (Formative) Evaluation, co-creation, science-practice engagement, climate change, criteria

THEORIES OF CHANGE FOR EVALUATING TRANSFORMATION-ORIENTED R&I PROGRAMMES: THE CASE OF THE 7TH ENERGY RESEARCH PROGRAMME IN GERMANY

MICHAEL DINGES, CHRISTIANE KERLEN, PETER KAUFMANN, ANNA WANG, KATHLEEN TOEPEL, JAKOB KOFLER, STEFAN MEYER AND HARALD WIESER

DOI: 10.22163/fteval.2022.542

ABSTRACT

This article investigates how Theories of Change for transformation-oriented R&I programmes can be designed to better grasp system transformation processes and thereby set the basis for a deeper understanding of transformative impact mechanisms and programme learning.

The analysis is based within the realm of the energy system, which is an area of specific concern for socio-technical transformation. It focuses on the "7th Energy Research Programme" (EFP) of the German Federal Ministry of Economic Affairs and Climate Action, which is the key R&I policy instrument contributing to the transformation of the energy system in Germany.

The article shows how a programme theory approach can be combined with multi-level perspective innovation system thinking and the concept of transformative outcomes to increase the evaluability of complex, transformation-oriented R&I programmes.

BACKGROUND

Within the realm of transformation-oriented policies, the energy system is an area of specific concern for a sustainability transformation of our society. The energy sector produces at least two-thirds of total greenhouse-gas (GHG) emissions (cf. Ritchie and Moser 2020). To contribute to the achievement of the goals set in the landmark Paris Agreement of the United Nations Framework Convention on Climate Change in 2015, European Union (EU) efforts and efforts of EU Member States set clear targets for realising the Energy Transition ("Energiewende").

The EU aims to be climate-neutral by 2050 – an economy with net-zero greenhouse gas emissions is at the heart of the European Green Deal and in line with the EU's commitment to global climate action under the Paris Agreement. For reaching this goal, primary energy consumption should be halved by 2050 compared with 2008 and a renewables ratio of 60% to gross final energy consumption should be achieved. As an intermediate step, the EU has set a common goal for the European energy transition: sun, wind, water and biomass are to cover 40 percent of the EU's electricity demand by 2030.

To contribute to this European end, the German government adopted the Climate Protection Plan 2050 to become largely greenhouse gas neutral by 2050. This goal was legally anchored in the Federal Climate Protection Act in 2019 (KSG 2019) and underpinned with measures in the Climate Protection Programme 2030. At the same time, it was decided to phase out coal by 2038 at the latest. In 2021, the federal government brought forward the long-term target from 2050 to 2045, aiming at an even faster energy transition.

R&I policy efforts at EU and EU Member States levels tackle the grand policy objectives for the Energy transition outlined above. At the European level, the Strategic Energy Technology (SET) Plan is a key endeavour for gearing R&I policies towards a sustainability transformation,[1] addressing the whole innovation system, and tackling both financing and the regulatory framework. Germany's "7th Energy Research Programme – Innovations for the Energy Transition" (EFP), is an outstanding example of governmental R&I programmes for a sustainability energy transition at the national level. The 7th EFP is assigned a key role in the German energy system transition by establishing a link between the long-term goals of the Federal Government and the time horizons of business technology research.

With the 7th EFP the federal government of Germany promotes research and development in the field of forward-looking energy technologies. It supports companies and research institutions to develop new technologies for energy supply, energy efficiency in sectors of

1 European Commission (2007), Strategic Energy Technology Plan. Retrieved March18, 2022, from https://ec.europa.eu/energy/topics/technology-and-innovation/strategic-energy-technology-plan_en#key-action-areas

consumption such as industry or housing, and system integration. Key new elements of the 7th EFP in comparison with its predecessors are:[2] 1) a stronger focus on technology and knowledge transfer including the introduction of a new instrument called "Living Labs for the Energy Transition" as a new programme pillar; 2) a broadening of the research spectrum that previously centred on individual technologies to encompass systemic and inter-systemic issues; and 3) a stronger focus on networking with international and European research.

The emphasis on the systemic character of transformation processes, an increased focus on cross-technology issues, system integration, and sector coupling play a central role in the programme. Furthermore, the embedding of individual technologies in overall societal trends and in the various sectors of energy generation and consumption are put focus on. In this way, government support for technology development and innovation is set to make a significant contribution to accelerate the transformation of the energy system, strengthening the industrial competitiveness and provide risk prevention for society as a whole.

RESEARCH QUESTION AND APPROACH

During early 2021, the four-year accompanying evaluation of the R&I funding measures and the accompanying measures of 7th Energy Research Programme have been initiated. The evaluation focuses on the non-nuclear research activities. The evaluation aims at developing analyses, reflections and recommendations as a basis for steering and continuous improvement of the programme ("programme learning") on the one hand, while also contributing to an assessment of programme effectiveness and impact on the other hand.

The evaluation commenced its work against the increasing need to frame the R&I programme evaluation in the energy system transformation context, and the key research questions of this article are:

- How can theories of change set the basis for an understanding of transformative impact mechanisms and programme learning?
- How can concepts of transitions in socio-technical systems extend theories of change to better capture transformation processes?

Our research is embedded in the ongoing planning process of the accompanying evaluation for the EFP. For establishing a theory of change of the programme, we investigate and test how a programme-theory based evaluation approach (Funnell and Rogers 2011; Rogers 2014) can be combined with 1) a multi-level perspective of system innovation (Geels et al. 2017), and 2) the concept of transformative outcomes (Ghosh et al. 2020, 2021).

Our research process includes the following key steps: First, we delineate the strategic and operational objectives of the programme including its design principles. Based upon this initial analysis of design characteristics and instrumental setting, we then show how the 7th EFP can be positioned in a multi-level perspective that grasps the manifold objectives and instruments of the programme and puts them in the context of the energy system transformation.

In our analysis, we then proceed to elaborate a theory of change for each instrument and identify main impact pathways that are intended to transform the energy system.

Based upon this analysis, we finally investigate whether and how the categories of transformative outcomes (see Ghosh et al. 2021) can be used to better understand the impact mechanisms of the programme and hence to increase its evaluability. Ghosh et al. show how transformation-oriented innovation programmes can shape ongoing transformation dynamics. Studying what types of Experimental Policy Engagements are most suitable for enabling transformation, the authors define three general spatially-bounded macro processes and sub-processes referred to as "transformative outcomes"[3] that policy actors can have some control over.

By assessing to what extent these processes can be observed in practice, they can be considered and reflexively addressed on an ongoing basis, e.g. in a formative evaluation, as a basis to continuously improve their policy engagements. Therefore, we research the correspondence of transformative outcomes and impact pathways within the overall programme theory.

OBJECTIVES AND INSTRUMENTS OF THE 7TH EFP

The 7th EFP has the ambition to contribute to the overarching energy and climate targets of the German federal government, aiming to be climate neutral by 2045, reduce carbon emissions by 88% in 2040 and by 65% in 2030 compared with 1990 (see KSG 2019).

The analysis of the programme documents and its regulations show that the programme is characterised by the following funding principles: 1) technology neutrality; 2) extension of project funding to include system integration & cross-system topics; 3) a focus on technology and innovation transfer & innovation-friendly framework conditions; and 4) strengthening international / European cooperation.

The programme and its instruments pursue the following strategic and operational objectives that are related to the energy transition challenges. The strategic objectives are:

- advancing the energy transition (develop holistic, innovative solutions & launching them quickly on the market; create environmentally-friendly, safe, and economic energy supply; activate innovation dynamics),
- strengthen the industrial competitiveness through modernisation, maintenance and development of competences, and creation of export opportunities of competitive technologies, and
- prevention of societal risks through a diversity of technological options.

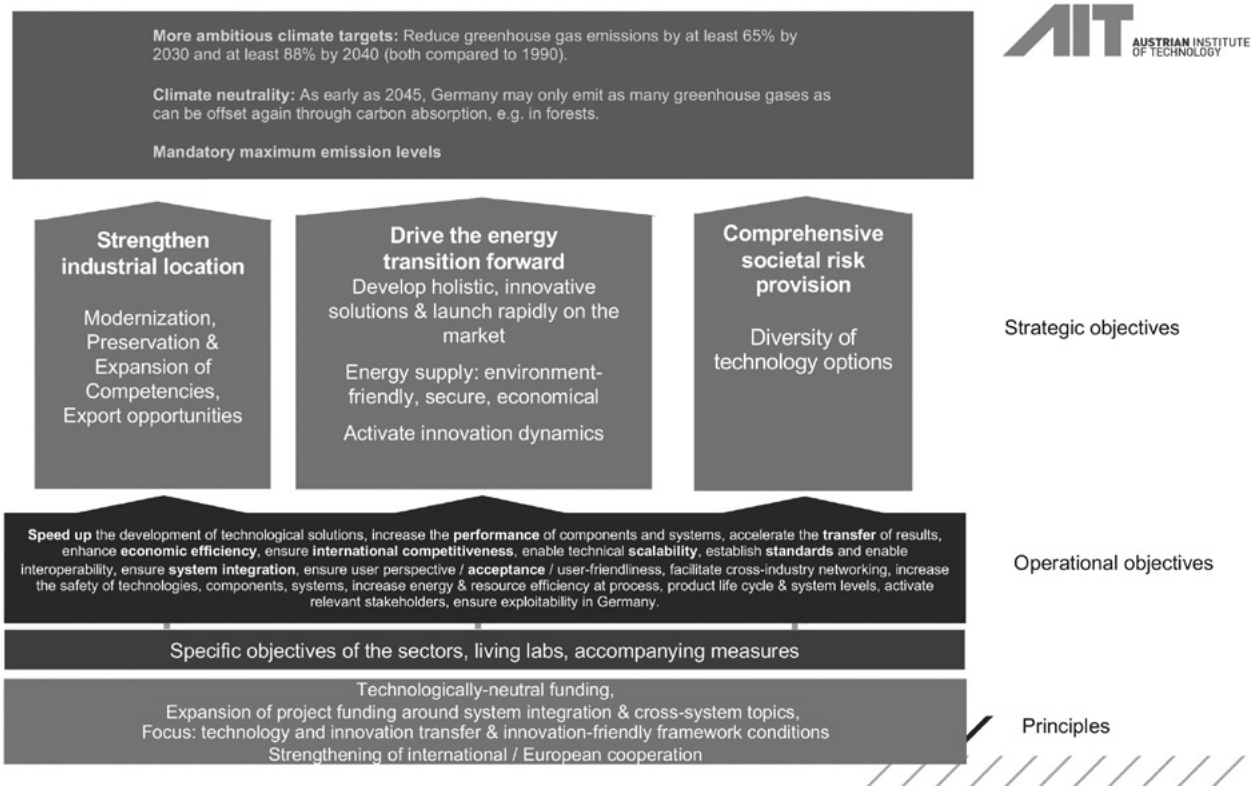
The operational targets are: 1) to develop technological solutions faster; 2) to increase performance of components and systems; 3) to

2 Bundesministerium für Wirtschaft und Energie (2018), 7th Energy Research Programme of the Federal Government. Retrieved March18, 2022, from https://www.bmwi.de/Redaktion/EN/Publikationen/Energie/7th-energy-research-programme-of-the-federal-government.pdf?__blob=publicationFile&v=5

3 From an evaluation perspective it seems confusing to refer to processes to enable change as outcomes, because outcomes are generally understood as changes in the status quo that result from an intervention. One may therefore consider "transformative outcomes" as introduced by Gosh et al. rather as "transformative mechanisms" inducing transformative change.

accelerate transfer of results; 4) to increase economic efficiency; 5) to ensure international competitiveness; 6) to enable technical scalability; 7) to establish standards and enable interoperability; 8) to ensure system integration; 9) to ensure user perspective / acceptance / user-friendliness; 10) to enable cross-sector networking; 11) to increase the safety of technologies, components and systems; 12) to increase energy & resource efficiency at the process, product life cycle and system levels; 13) to activate relevant actors; and 14) to ensure exploitability in Germany.

Figure 1: Objectives of the 7th Energy Research programme
Source: Own illustration



The 7th EFP tackles the energy system transformation through three different types of instruments: 1) R&I projects; 2) Living Labs, and 3) Accompanying Measures. The three instruments are collectively geared towards the supply of new technologies (technology push), the speeding up of new knowledge and technology transfer (demand pulls), and system development efforts. They target practices within:

1) the renewable energy supply system (solar energy, geothermal energy, wind energy, biomass from plants, and hydropower) and their system integration
2) the energy consumption sectors (e.g. industry, transport, buildings and neighbourhoods), and
3) the development of green substitutes for carbon-based technologies, e.g. fuel cell technologies.

THE 7TH EFP FROM A MULTI-LEVEL PERSPECTIVE

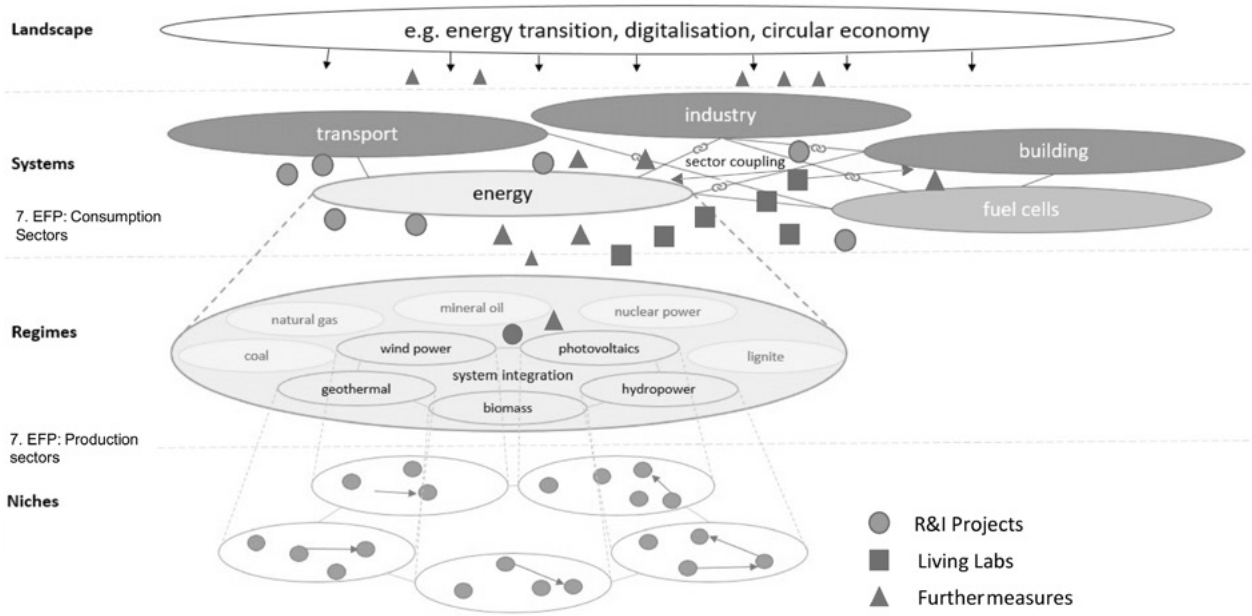
For the analysis of the 7th EFP we make use of the multi-level perspective of system innovation. The multi-level perspective was designed as a broad heuristic to capture transitions in different socio-technical systems such as mobility or food (EEA 2018; Geels et al. 2017). The basic idea is that due to existing path dependencies, dominant regimes (e.g. energy production from fossil fuels) can only be changed through profound technological and social measures that simultaneously destabilise the regimes and generate spaces for radically different solutions. In its standard form, the multi-level perspective differentiates between three levels (Geels 2006):

- The macro level is the socio-technical landscape, i.e. the wider exogenous environment that affects socio-technical development (e.g. globalisation, environmental challenges, policy framework etc.);
- The meso-level is formed by socio-technical regimes providing the rule-set or grammar embedded in a complex of engineering practices, production process technologies, product characteristics, skills and procedures, ways of handling relevant artefacts and persons, ways of defining problems; all of them embedded in institutions and infrastructures (Rip and Kemp 1998);
- The micro level is formed by technological niches that emerge in "protected spaces", which act as "incubation rooms" for radical novelties (Schot 1998, Kemp et al. 1998), to shield them from mainstream market selection.

Originally developed for the analysis of individual regimes, the multi-level perspective has since been extended to make interactions of different regimes and systems tangible (see Rosenbloom 2020). Such a broader perspective makes it possible to reflect and systematise the various technical levels of analysis ("fachliche Betrachtungsebenen") that informed the design of the 7th EFP: consumption sectors, energy production, system integration, cross-system research topics, and the accompanying measures (i.e. establishment and support of sector networks, accompanying studies, public relations, research communication at programme level).

Figure 2 illustrates how the instrumental setting of the 7th EFP and its structural composition can be analysed and portrayed following the heuristic of the multi-level perspective. The figure shows that the different intervention mechanisms of the programme tackle to varying degrees: 1) different levels of the socio-technical system; and 2) different sub-regimes (energy production sectors) and systems (energy consumption sectors). Therein, the impact creation process of the programme can be understood as a process of mutual alignment of developments at different levels.

Figure 2: Technical Levels of Analysis, Sectors, and instruments of the 7th EFP in a Multi-Level Perspective
Source: Own illustration



For the operationalisation of the evaluation, the following lessons can be drawn from the positioning of the programme in a multi-level perspective:

First, we acknowledge that the application of the multi-level perspective allows to facilitate a close correspondence between the programme theory (objectives and intervention mechanisms) and the perspectives of programme managers. As such, it is an important attribute that improves the evaluation's transparency and facilitates the generation of a shared understanding of the programme.

Second, we see that the multi-level perspective allows integrating considerations of production and consumption sectors. The programme addresses a number of specific consumption and production sectors are. Each sector has specific targets which are to be achieved next to achieving the overarching operative programme objectives. For example, the analysis of objectives of the specific sectors of production and consumption shows that sector-specific targets encompass: 1) specific goals for existing technologies (e.g. new processes, products, applications); 2) energy and cost efficiency-related goals; 3) goals related to the substitution of specific materials or technologies; and 4) goals related to the system integration of new technologies.

Third, we see that the sector-specific targets of the R&I programme reflect challenges at the regime and system level, which the R&I instruments of the programme shall address. For the operationalisation of the empirical evaluation, they allow to define sector-specific hypotheses concerning the relevance and coherence of objectives, the appropriateness of challenges addressed by R&I portfolios, and the required characteristics of actors involved in the process.

Finally, stemming from the multi-level perspective, a decisive factor, and a challenge for the accompanying evaluation of the 7th EFP, is to analyse to what extent the measures of the 7th EFP correspond to the developments at the regime and landscape level. The multi-level perspective is a heuristic indicating that the 7th EFP does not exist in isolation but is embedded in a socio-technical system that is not only shaped by R&I initiatives alone but also demand-side policies, regulatory policies, socio-economic trends and market structures at the regime level, that shape the conditions for knowledge diffusion, societal acceptance of transition process, speed and uptake of new solutions.

However, the multi-level perspective does not specify which landscape and regime dynamics need to be taken into account. We therefore suggest performing a STEEP analysis in evaluations to create a structured overview of the factors that may spur or impede the progress of technology uptake, the types of topics and research challenges that should be addressed, and the stakeholders that should be considered in future activities.

The multi-level perspective also draws attention to the fact that R&I funding is only one measure within the toolbox of innovation policy geared at enabling transformational change. Compared to other innovation-oriented instruments like regulatory reform or financial incentives for the uptake of new technologies, R&I funding is confronted with the challenge of (a) high uncertainty about success, (b) long time-spans before results have a tangible effect, and (c) small budgets compared to other instruments using public spending, and (d) dependencies of other related policies such as regulatory reforms and sector specific policies.

Figure 3: STEEP analysis – System trends exerting impact on the EFP
Source: Own compilation

Social	Technological	Economic	Environmental	Political & Legal
• Societal acceptance of the energy transition • Social disparities • Poverty risk • Housing/Living standards • Public health disparities • Discourse about distributive justice • Civil society movements • Energy consumption patterns • Position towards nuclear energy	• Low-carbon hydrogen and hydrogen-based fuels • Carbon capture technologies • Innovation speed & upscaling • Cross-cutting technology development • Digitisation • Sector-Coupling • Decentralisation and flexibility of the electricity infrastructure • Energy efficiency technologies for high-carbon industries • International cooperation • Costs for alternative energy technologies	• Clean electricity roll-out • Diversity in energy suppliers • Energy technology start-ups & supply market development • Energy market integration: international, national • Energy costs • Energy imports • Turnover of renewables, and energy related technologies	• Climate change: global warming • Extreme weather events: constant rain, snowfall, thunderstorm, hail drought, strong winds and storms, floods, heat waves	• Active design of structural change: Economic stimulus /recovery package • German Development and Resilience Plan (DARP) • Regulation: End Coal-Fired Power Generation • Climate-friendly mobility act • Hydrogen Strategy • Offshore Wind Energy Act • Emission trading and CO2 prices • International conflicts & availability of raw materials • Energy taxes • Energy market regulation • Sustainability Investments Regulations (EU-Taxonomy, Corporate Sustainability Directive)

PATHWAYS TO IMPACT

Departing from the investigation of the instrumental setting of the programme from this multi-level perspective, we can further dive into the investigation how the programme seeks to achieve its objectives and eventually contribute to system transformation.

The programme theory approach (see Rogers 2014) connects the underlying rationales of a programme (a specific challenge to be addressed), with an overall roadmap on how specific activities are expected to produce immediate outputs connected to outcomes/intermediate impacts and eventually the realisation of the objectives.

As the 7th EFP tackles the energy system transformation through three different types of instruments (R&I Projects, Living Labs for the Energy Transition, Accompanying Measures), we developed instrument-specific theories of change based on an analysis of 1) programme documentations and interviews with representatives from 1) the responsible federal ministry, 2) the managing agency, and 3) R&I actors engaged in the programme.

While sharing this overall orientation, each instrument pursues a set of different activities and rests upon specific pathways to impact, which aim to mutually reinforce each other. (Figure 5 – full theory of change for R&I projects):

- The **R&I-projects** addressing single technologies follow impact pathways through the generation of new knowledge, qualification, technological innovation, and transfer of pilot demonstrations into business practices. Transdisciplinary research projects focus on system development through focusing on cross-systemic issues of the energy transition.
- The **Living Labs** projects follow pathways of developing complete system solutions for the whole energy sector. Testing and piloting of solutions in real world environments, networking of main energy system actors for collectively avoiding CO2 emissions are key impact pathways.
- The **Accompanying Measures** are collectively geared towards accelerating the creation of impact at the regime level through synthesising knowledge, increasing circulation and transfer in research networks, enabling new partnerships, enhancing qualification and increasing transparency.

Figure 4: Instrument-specific Activities and Impact Pathways
Source: Own compilation

R&I Projects, Pilots & Demonstrators		Living Labs		Accompanying Measures	
Activities	**Pathways**	**Activities**	**Pathways**	**Activities**	**Pathways**
Individual R&I projects on single technologies Collaborative R&I projects on single technologies Pilot Projects & Demonstration projects Transdisciplinary research projects on systemic and cross-systemic issues of the energy transition	Knowledge creation & capacity development Networking Economic valorisation Transfer System development	Collaborative R&I in Living Lab contexts related to: • Digitalisation, ICT development • Reflection of experimentation clauses • Developing and building industrial plants • Test / pilot operation / demonstration • Supplementary R&D on individual issues Living Lab Coordination	Innovation Upscaling Avoiding CO2 emissions Diffusion	Establishment and support for Energy Transition Research & Innovation Platform and Research Networks Accompanying research and studies Research Communication Public Relations at programme level	Synthesizing knowledge Knowledge circulation & and transfer Enabling cooperation Increasing qualification Increasing transparency

Figure 6 illustrates the complete theory of change including the main impact pathways of 1) knowledge generation, 2) network creation, 3) innovation, 4) transfer, and 5) system development. The theory of change following impact pathways (see Douthwaite et al., 2003) is a model describing how the programme and activities therein seek to achieve impact and allows for a better attribution of programme activities to impacts achieved. The theory of change has been elaborated in an iterative process of documentary analysis of programme instruments, interviews of project participants, and reflections with programme authorities (ministries and programme management).

The theory of change is a model that is capable of explicitly illustrating causal hypothesis of programme interventions and aspired impacts of new programmes (Balthasar and Fässler, 2017). Taking an intervention-based perspective with a focus on programme actors, it illustrates a sequence of conditions that must be achieved for a problem to be solved (see Clark and Anderson, 2004). The theory of change is necessarily a reduction of complexity and therefore has tendencies to omit context consideration.

In our analysis of the 7th EFP, the multi-level perspective provides means to better understand the specific intervention mechanisms of the programme and allows to pose a number of evaluative questions for the collection of empirical data, relating to key impact pathways and the socio-technical innovation system. We will answer the "classic" evaluation questions about effectiveness, efficiency, relevance, coherence etc. in the evaluation, but the MULTI-LEVEL PERSPECTIVE leads to some additional questions and specific perspectives that we want to address. Examples of questions relating to key impact pathways in the context of the multi-level perspective are:

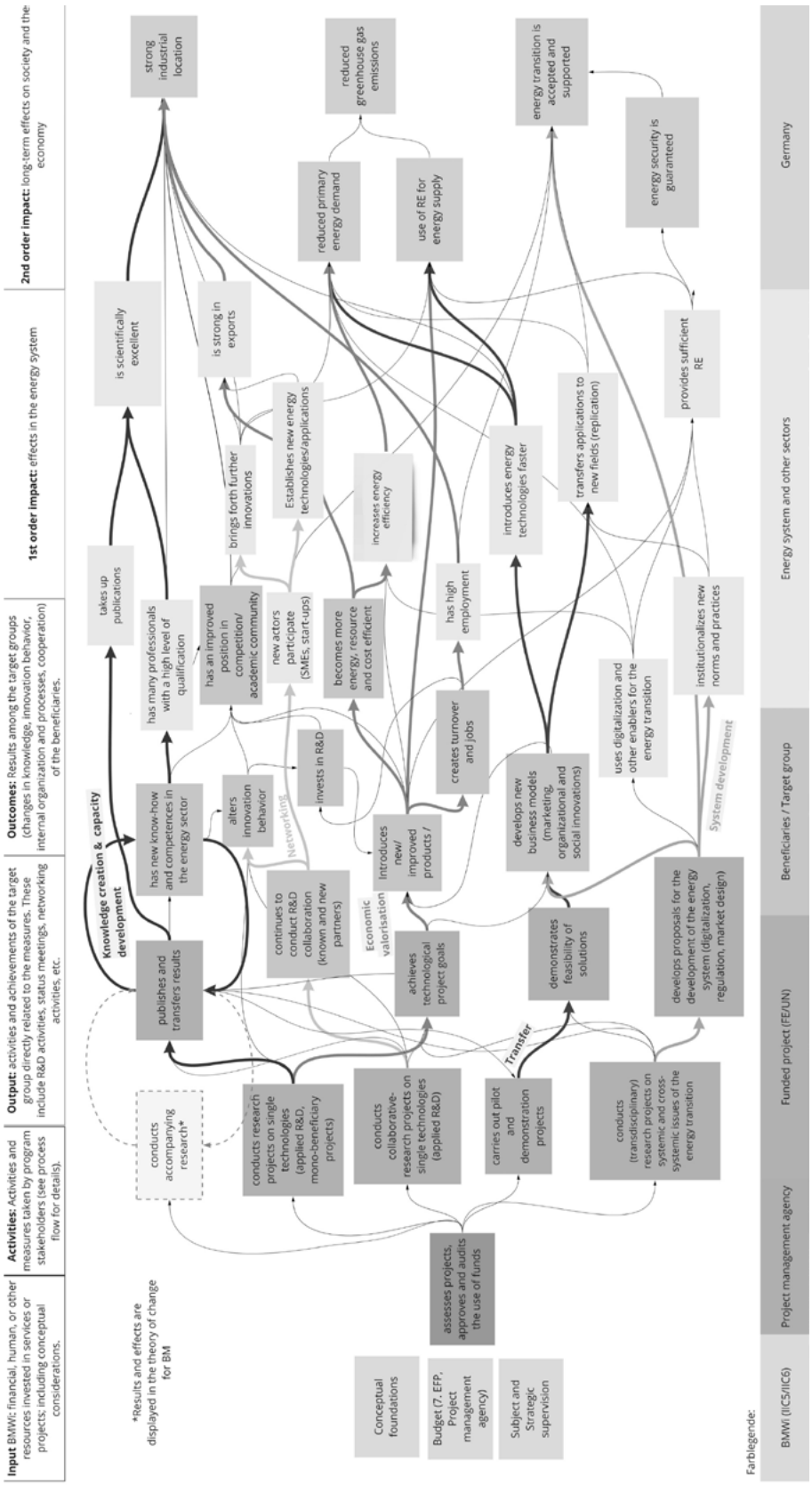
PATHWAY 1: THE KNOWLEDGE CREATION AND CAPACITY BUILDING PROCESS

- Which actors are performing the research and development work in the programme? How are they anchored in the socio-technical innovation system?
- To which extent is there an active communication and dissemination of learnings and outcomes of innovative practices of research results?
- Is the knowledge being primarily generated for the right stakeholders? Those who could bring it to application??
- Does capacity building encompass only existent regimes or does it prepare for niches and their training and qualification needs?
- Are skills and procedures, ways of working, rules and regulations objects of research? How is this knowledge being transferred?

PATHWAY 2: NETWORK CREATION

- Which actors are actively involved in the knowledge generation process of the projects? What is their role in the socio-technical innovation system?
- Which activities are being performed that enable knowledge exchange between owners, producers, and researchers on a regular basis?
- To which extent is there collaboration across organisational boundaries for certain focus areas/topics in the projects?

Figure 5: Theory of Change "R&I Projects"
Source: Own compilation



- Are actors involved or being addressed that are of particular importance for the transformation of the energy sector? (E.g. energy communities, the again increasing number of municipal energy providers/utilities, IT companies, start-ups).
- How are new actors involved in the process of knowledge generation and diffusion that can promote a transformation of the energy system?
- How do incumbent regime actors position themselves vis-à-vis transformation processes in the socio-technical innovation system? What role do incumbent regime actors play in knowledge diffusion and transfer of the programme?

PATHWAY 3: INNOVATION

- Do the research topics addressed contribute to the creation of energy technologies that are new to the market? Do they potentially open up new niches for the sustainable transformation of the energy system (e.g. technologies for better coping with the decentralisation of energy supply)?
- What contribution does the research make to increasing the market penetration of niche technologies (for example regarding cost reduction of renewable energies/applications)?
- What are the economic application potentials of the new technologies?
- Which measures within the programme/the project are being taken to increase chances of market introduction or market penetration?
- What are the characteristics of the markets targeted by the new technologies? What is the positioning of incumbent regime actors? Are there new market entrants?
- What is the level of entrepreneurial activities in the specific socio-technical innovation system?
- Which elements of the socio-technical innovation system increase or decrease possibilities for market introduction and sustainable market penetration of newly developed technologies?

PATHWAY 4: TRANSFER

- What is the technological maturity of the research and development subjects? Can their application be demonstrated successfully?
- Which activities are being performed for developing new business models that reach new users in different market, and/or shorten supply chains, making the innovation accessible to a broader audience?
- To what extent are there shared goals that facilitate successful interaction and learning?
- Which activities are being performed that enable cross-sectoral exchange between different type of actors?
- Are there any complementary actions in addition to R&D subsidies within the programme/outside the programme that support speeding up market formation and market penetration?
- Which regulatory aspects need to be considered/changed for the introduction/increased market penetration of new technologies?
- Are the research activities leading to good practices and to the creation of guidelines so that innovative practices can be implemented in a different context, such as a different location?

PATHWAY 5: SYSTEM DEVELOPMENT

- What is the extent of transdisciplinary research dealing with socio-technical system development cross-cutting research topics?
- Are research results providing evidence for high-level policy making?
- Is there a promotion for the formalisation of ideas and new practices by incumbent actors (local policy bodies, industry associations, etc.)?
- Do the research results have an impact on changes in standards / norms / other regulatory aspects (e.g. providing guidance in the development of new regulations and norms that support an innovation)?
- To which extent are suggestions for market development and regulatory frameworks being taken up in the policy discourse?

IMPACT PATHWAYS AND TRANSFORMATIVE OUTCOMES OF THE 7TH EFP

For positioning the impact creation process in the energy transition policy context, we analyse to what extent the programme and its instruments target the energy transition. Raven et al (2016) accentuate that this needs to be realised through strategic niche support. Furthermore, this also requires deconstructing and working against harmful regimes (Turnheim and Geels 2012; Kiviima and Kern 2016). The above introduced concept of "transformative outcomes" identifies three general spatially-bounded macro processes that actors can have control over: (1) building or nurturing niches; (2) expanding and mainstreaming niches, and (3) opening up and unlocking regimes. In each of these three macro-processes, four sub-processes were identified that actors (e.g. programme owners/managers and project leaders) can have control over. The transformative outcomes are not in any particular order and can "co-evolve through time and space". They provide more granular categories which specify important leverage points for niche development and regime destabilisation.

While the transformative outcomes are described in detail in Ghosh et al. (2021) we focus here on the question to which extent the instruments and identified impact pathways of the programme correspond with the transformative outcomes outlined therein.

Figure 6: Impact Pathways and Transformative outcomes of the 7th EFP
Source: Own compilation elaborating on Gosh et al. (2020)

R&I Projects	Living Labs	Accompanying Measures	Transformative Outcomes
Impact Pathways			
Building and Nurturing Niches			
Knowledge Generation	Innovation		**Shielding**: protecting new and more sustainable practices from external influences and helping them grow
		Knowledge Circulation & Transfer	**Learning**: providing regular opportunities for discussing experiences, obstacles and needs related to a new practice as well as challenging related values and assumptions that people might have
Network Creation		Enabling cooperation	**Networking**: protecting and progressing new practices by gaining interest of more people and creating connections between them
System development		Synthesising Knowledge	**Navigating expectations**: navigating and converging expectations of different actors the legitimacy of new practices is developed and their potential explored
Expanding and mainstreaming niches			
Economic valorisation	Upscaling		**Upscaling**: conducting deliberate action to get more users involved into new and more sustainable practices
Transfer	Diffusion	Increasing qualification	**Replicating**: transferring the new and more sustainable practices to another location
		Enabling cooperation	**Circulating**: exchange of knowledge, ideas and resources between multiple related alternative practices
	Diffusion		**Institutionalising**: turning new and more sustainable practices into more permanent and more widely available ones
Opening up and unlocking regimes			
			De-aligning and destabilising regimes: disrupting and weakening dominant practices. This can be done by changing one of the dominant dimensions for example through the introduction of new policies
System development	Avoiding CO2 emissions	Transparency	**Unlearning and deep learning of regime actors**: dominant actors question their assumptions and change their view on the potential of new and more sustainable practices and the ability of the dominant practice to respond to threats and opportunities, such as climate change and digitalisation
Network creation		Enabling cooperation	**Strengthening regime-niche interactions**: Frequency and quality of interactions between empowered actors from the niche and the regime on a non-competitive basis
		Synthesising knowledge	**Changing perceptions of landscape pressures**: dominant actors to reach the point of view that immediate action is warranted, and new emerging more sustainable narratives need to be promoted

In the context of the 7th EFP we can assume that shielding of R&D activities is a key function being provided through direct R&D funding in the R&I projects and Living Labs that address all innovations necessary for system innovation (e.g. technology, organisational, business models, etc.). The R&D funding of the programme provides a protected space for developing new ideas that aim to spur the technological advancement of the energy system. The living labs also provide research infrastructures and a targeted R&I portfolio that support experimentation with niche technologies and niche actors.

Network creation through R&I projects and specific instruments of the "Accompanying Measures" of the 7th EFP are supposed to gather research, user and policy communities and facilitate collective **learning**

and **networking**. While the collaborative R&I projects build networks starting from the project level, nine "Energy Research Networks"[4] represent the broad research landscape on the topics of bioenergy, buildings and neighbourhoods, renewable energies, flexible energy conversion, industry and commerce, electricity grids, start-ups, system analysis and hydrogen. The networks are supposed to be dialogue-oriented forums for exchange between research, politics and industry and offer space for a self-organised process of their members. In terms of transformative outcomes illustrated by Ghosh et al. (2021), they provide room for synthesizing knowledge, discussion of alternative ideas, reflection and learning.

It can also be expected that **navigating expectations** is a deliberate result of the Accompanying Measures and the R&I projects of the 7th EFP. Within the instrument of R&I projects, system development is being promoted through support of R&I focussing on cross-system topics and system integration. As a result, contributions to the development of standards, norms and other regulatory aspects, as well as high-level policy making should arise. Within the Accompanying Measures, the "Energy Transition Platform for Research & Innovation", which acts as an advisory body for the Federal Ministry for Economic Affairs and Climate Action has the function to facilitate dialogue on the strategic direction of energy research policy with national stakeholders from politics, business, science and society. A main task of the Energy Transition Platform for R&I is to synthesize the collective knowledge gathered in the research networks.

Expanding and mainstreaming niches can mostly be related to the R&I activities of the "Living Labs". The process of upscaling in the living labs aims to increase the reference capability of novel technological solutions, which should turn into novel standard operations at the regime level and contribute to cost-reductions of these novel technologies. Diffusion is linked to the process of introducing system solutions, building and applying blue-prints and the diffusions of new processes/standard practices at the level of the energy system and other sectors. While these outcomes are also included somewhat in the R&I projects, they are more explicitly formulated and aspired in the Living Lab concept of the programme.

In the context of expanding and mainstreaming niches, the accompanying measures may have an amplifying function, as they aim to provide means for collective exchange of knowledge, ideas and resources between multiple related alternative practices in a self-organised manner.

When it comes to **opening-up and unlocking regimes** one should primarily be aware that R&I policies and instruments might not be the most powerful tool to rely upon. An introduction and implementation of new regulatory policies, changes in fiscal policies (prices/taxation) may challenge and trigger the search for new solutions much more effectively than technologically open R&I programmes.

Nevertheless, for the 7th EFP the pathways of Network Creation (R&I projects) and Enabling Cooperation (Accompanying Measures) are presumably also functions of the EFP networks for **strengthening regime-niche interactions**, whereas Living Labs seek to deeply change the path of existing regimes through CO2 avoidance and sectoral diffusion of new solutions. Furthermore, the pathway of **synthesising knowledge** may contribute to alter perceptions of main regime actors concerning landscape pressures and start to pursue new pathways.

Just from this exercise of describing the relationship between instruments and pathways of the 7th EFP and the concept of "transformative outcomes", it becomes clear that the programme has not only a transformative ambition but might be able to contribute to change existent socio-technical innovation systems. For advancing the operationalisation of the evaluation, the exercise shows that the bottom-up created Theory of Change of EFP instruments including their pathways of impacts can be related to transformative outcomes, which allows to better tailor the empirical evaluation design towards the relevance, coherence and effectiveness of the programme in terms of its contribution to the energy transition.

REFLECTIONS

Through establishing an integrated programme theory for the 7th EFP, we show how predominantly linear theories of change can be enhanced by integrating a multi-level perspective and transformative outcomes. The emerging programme theory reflects the need to develop formative and embedded monitoring and evaluation of transformation-oriented R&I programmes, embedded in a multi-level perspective.

Putting the evaluation of the 7th Energy Research Programme in the multi-level perspective facilitates 1) taking a more dynamic perspective on the intervention mechanisms of the evaluation object and 2) better integrating external factors at the regime and landscape level that exert influence on the effectiveness of the programme. Positioning the programme in the multi-level-perspective shows that building and nurturing niches, with the ambition to replicate and upscale technological system innovations at the regime levels of energy production and consumption is the main impact mechanism of the instruments R&I projects and Living Labs, while certain parts of these instruments also cover developments at the landscape level (i.e. through transdisciplinary research projects and system analysis of the energy transition process), and the interaction between different consumption and production sectors. The various accompanying measures of the programme aim to contribute to synthesizing collective knowledge, niche-regime interactions beyond the project level and navigating expectations.

While the theory of change can be created in a bottom-up manner, based on programme documentation, views and perspectives of programme management and project participants, the multi-level perspective is a heuristic that allows to frame hypotheses and questions concerning the impact creation process, and facilitate programme learning. For instance, for considering the contribution of the EFP to opening-up and unlocking regimes, the frequency and quality of interactions in the socio-technical innovation systems needs to be explored by the evaluation as well as the changing perceptions and actions of actors in the socio-technical system. In the case of regime-niche interactions, the evaluation will also have to consider path-dependencies and rigidities of incumbent regimes causing a lock-in in existing trajectories. For example, as regards the energy transition Ghosh et al. (2021) warn that even when alternatives are proposed by regime actors, they tend to reaffirm the architecture of the system as it is.

4 See: Projektträger Jülich (n.d.), Forschungsnetzwerke Energie, retrieved March 18, 2022, from https://www.forschungsnetzwerke-energie.de/forschungsnetzwerke-energie

While the combination of the multi-level perspective with an input-output-outcome-impact model at an aggregate level increases accountability, it remains a key challenge to define indicators that reflect the complexity of transformation processes on the one hand, while specifically detailing the contribution of a programme towards these processes on the other. As the programme theory delineates the main pathways to impact of the programme and considers external influential factors as well, it should not only allow for deep learning loops for programme owners and actors in the programme but also contribute to an enhanced evaluability of transformation-oriented R&I programmes. In this regard, our analysis has shown that the inductively generated impact pathways from the theory of change can be looked at through the lens of transformative outcomes. An additional advantage of using a perspective which has been derived from the literature is better comparability with other transformative RTI programmes.

REFERENCES

Balthasar, A., & Fässler, S. (2017). Wirkungsmodelle: Ursprung, Erarbeitungsprozess, Möglichkeiten und Grenzen.

BMWI (2020), 7th Energy Research Programme of the Federal Government, Berlin.

Clark, H., & Anderson, A. A. (2004, November). Theories of change and logic models: Telling them apart. In American Evaluation Association Conference. Atlanta: GA.

Douthwaite, B., Kuby, T., van de Fliert, E., & Schulz, S. (2003). Impact pathway evaluation: an approach for achieving and attributing impact in complex systems. Agricultural systems, 78(2), 243-265.

EEA (2018). Perspectives on transitions to sustainability. European Environmental Agency, Report 25/2017, Luxembourg.

Funnell, Sue C., Rogers, Patricia J. (2011): Purposeful Program Theory: Effective Use of Theories of Change and Logic Models.

Geels, F. W. (2002). Technological transitions as evolutionary reconfiguration processes: a multi-level perspective and a case study. Research Policy, 31, 1257-1274.

Geels, F. W., & Schot, J. (2007). Typology of sociotechnical transition pathways. Research Policy, 36, 399-417.

Geels, F. W., Kern, F., Fuchs, G., Hinderer, N., Kungl, G., Mylan, J., Neukrich, M., & Wassermann, S. (2016). The enactment of socio-technical transition pathways: A reformulated typology and a comparative multi-level analysis of the German and UK low-carbon electricity transitions (1990-2014). Research Policy, 45, 896-913.

Geels, F. W., Turnheim, B., Asquith, M., Kern, F., & Kivimaa, P. (2019). Sustainability transitions: policy and practice. Retrieved March 18, 2022, from https://www.eea.europa.eu/publications/sustainability-transitions-policy-and-practice

Geels, F. W., Sovacool, B. K., Schwanen, T., & Sorrell, S. (2017). Socio-technical transitions for deep decarbonization. Science, 357, 1242-1244.

Ghosh, B., Kivimaa, P., Ramirez, M., Schot, J., Torrens, J., 2020. Transformative Outcomes: Assessing and reorienting experimentation with transformative innovation policy Transformative outcomes, TIPC Working Paper, TIPCWP 2020-02. Retrieved March 18, 2022, from http://www.tipconsortium.net/wp-content/uploads/2020/07/Transformation-outcomes-TIPC-working-paper.pdf

Ghosh, B., Kivimaa, P., Ramirez, M., Schot, J., & Torrens, J. (2021). Transformative outcomes: assessing and reorienting experimentation with transformative innovation policy. Science and Public Policy, 48(5), 739-756.

International Energy Agency (2021). Global Energy Review 2021, Paris. Kivimaa, P., & Kern, F. (2016). Creative destruction or mere niche support? Innovation policy mixes for sustainability transitions. Research Policy, 45(1), 205-217.

KSG (2019), Bundes-Klimaschutzgesetz, last revision 31. August 2021 (Art. 2 G, 18. August 2021). Retrieved March 18, 2022, from https://www.gesetze-im-internet.de/ksg/index.html.

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.

Larrue, P. (2021), "The design and implementation of mission-oriented innovation policies: A new systemic policy approach to address societal challenges", OECD Science, Technology and Industry Policy Papers, No. 100, OECD Publishing, Paris.

Larsen, H. (2019): Capabilities, Networks and Directionality. Innovation Policy for the Sustainable Development Goals, PHD thesis Imperial College, London.

Lundin, N., & Serger Schwaag, S. (2018). Agenda 2030 and A Transformative Innovation Policy–Conceptualizing and experimenting with transformative changes towards sustainability. Work in process WP, 1.

Molas-Gallart, J., Boni, A., Schot, J., & Giachi, S. (2020). A Formative Approach to the Evaluation of Transformative Innovation Policy. Retrieved March 18, 2022, from http://www.tipconsortium.net/

Raven, R., Kern, F., Verhees, B., & Smith, A. (2016). Niche construction and empowerment through socio-political work. A meta-analysis of six low-carbon technology cases. Environmental Innovation and Societal Transitions, 18, 164-180.

Rip, A., & Kemp, R. (1998). Technological change. Human choice and climate change, 2(2), 327-399.

Ritchie, H., & Roser, M. (2020). CO^2 and greenhouse gas emissions. Our world in data.

Rogers, P. (2014). Theory of change. Methodological briefs: impact evaluation, 2(16), 1-14.

Rosenbloom, D. (2020). Engaging with multi-system interactions in sustainability transitions: A comment on the transitions research agenda. Environmental Innovation and Societal Transitions, 34, 336-340.

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.

Turnheim, B., & Geels, F. W. (2012). Regime destabilisation as the flipside of energy transitions: Lessons from the history of the British coal industry (1913–1997). Energy policy, 50, 35-49.

AUTHORS

MICHAEL DINGES
Austrian Institute of Technology
Giefinggasse 4, 1210 Wien
E: michael.dinges@ait.ac.at

CHRISTIANE KERLEN
Kerlen Evaluation
E: christiane.kerlen@kerlen-evaluation.de

PETER KAUFMANN
KMU Forschung Austria / Austrian Institute for SME Research
E: p.kaufmann@kmuforschung.ac.at

ANNA WANG
AIT Austrian Institute of Technology
E: anna.wang@ait.ac.at

KATHLEEN TOEPEL
Kerlen Evaluation
E: kathleen.toepel@kerlen-evaluation.de

JAKOB KOFLER
KMU Forschung Austria / Austrian Institute for SME Research
E: j.kofler@kmuforschung.ac.at

STEFAN MEYER
Kerlen Evaluation
E: stefan.meyer@kerlen-evaluation.de

HARALD WIESER
KMU Forschung Austria / Austrian Institute for SME Research
E: h.wieser@kmuforschung.ac.at

KEYWORDS: system transformation, programme theory, energy system, energy transition, transformative outcomes

DRIVING THE INNOVATION PROCESS BY CONNECTING REGIONAL KNOWLEDGE BASES TO LOCAL NEEDS

SONIA DANIELA MENA JARA, INGEBORG MEIJER, GASTON HEIMERIKS AND TIM WILLEMSE

DOI:10.22163/fteval.2022.543

ABSTRACT

Novel approaches are needed to support the creation of more open, inclusive, and self-sustaining R&I ecosystems in healthcare. This study analysed 3 European regions (Murcia ES), (Örebro SE), and (Republic of Cyprus CY), incorporating complementary approaches from Responsible Research and Innovation (RRI) and Research and Innovation Smart Specialisation Strategy (RIS3). The exercise entailed the identification of healthcare and innovation stakeholders and the characterisation of the policy landscape in each territory. Moreover, the strengths of the regional knowledge base was analysed by measuring the Revealed Comparative Advantage (RCA) indicator based on relatedness measurement, and by using micro-level fields analyses of scientific publications. This methodology allowed us to identify the fields and topics (strengths) that provide opportunities for innovation processes. Additional identification of social needs in the three territories showed profound differences regarding the alignment of the selected needs with respect to the regions' capabilities. The results suggest that a timely direct interaction with territorial stakeholders can help in selecting the most promising innovation priorities that are based on local needs and knowledge. The process of interaction requires early engagement to support territorial ownership and is further reinforced by RRI policies in place.

THEORETICAL BACKGROUND

In this paper we address the question whether we can articulate an approach to regional science and innovation strategies that not only promote smart (i.e., competitive) but also inclusive and sustainable regional economic development (i.e., responsible research and innovation). The approach we present emphasises co-creation processes in the regions where the entrepreneurial discovery process is taking place by encouraging the participation of a diverse set of actors. It encourages a bottom-up process towards the definition of societal expectation and local needs in the regional context. This process is performed jointly with the identification of the regional capabilities and skills (strengths) founded in the regional knowledge base; the analysis of specialised knowledge aligned to the identified needs can assist a successful innovation process.

Our objective is to develop an approach to assist policymakers and other stakeholders in designing and implementing Research and Innovation Strategies for Smart Specialisation (RIS3). By combining information on the relative strength of regional knowledge production activities with information about regional stakeholders, local needs, and policies, we can specify priorities that can help to maximise the regional development potentials.

How can smart specialisation be an effective tool to help regions discover new opportunities for more sustainable and inclusive societies? This paper shares outcomes of the analysis framed under the ongoing EU-funded project CHERRIES (*Constructing Healthcare Environments through Responsible Research Innovation and Entrepreneurship Strategies*), which strives to create more open, inclusive, and self-sustaining R&I ecosystems by enabling RRI policy experiments in the healthcare sector in three European territories – in Murcia (ES), Örebro (SE), and the Republic of Cyprus (CY). The project activities encompass 1) an initial stakeholder and policy mapping exercise, 2) the definition of regional R&I capabilities or strengths, 3) the identification of local demands through a call for needs-process, 4) the definition and implementation of local innovation pilots based on delineated needs and capabilities, and 5) the formulation of policy recommendations in the territorial context of each regional healthcare and innovation system. In this article we address activities 1, 2 and 3.

The underlying rationale behind the smart specialisation approach is that by concentrating knowledge resources and linking them to a limited number of priority economic activities, regions can become — and remain — competitive in the global economy (European Commission, 2012). The priority fields for each EU region are set in an entrepreneurial discovery process (EDP) by regional actors (Foray *et al.*, 2011; OECD, 2013). The RIS3 approaches the EDP with the idea of societal engagement in the form of participatory public-private dialogue. This collaborative model tries to create alignment between regional capabilities and regional policy by enabling regions to prioritize domains seen as important (Foray, 2016). The Responsible Research, Innovation, and Entrepreneurship Strategies need to build on existing strengths of a region ('smart specialisation') and should involve reflection on local values and needs (European Commission, 2014). This principle presents a knowledge gap for policymakers and other regional stakeholders.

Shaping the territorial dimension of science and innovation policies for inclusive and sustainable growth requires the understanding of the territorial diversity, opportunities, and constraints in knowledge developments of different places to maximise their potentials. The current scientific portfolio of a region influences the capacity to innovate (Heimeriks

et al., 2019). Just as regions differ in size and wealth, they also vary in the diversity and complexity of their knowledge base. Especially large, metropolitan regions are capable of contributing to a wide range of fields (Nomaler et al., 2014). In contrast, the ability of regions to diversify into new fields of knowledge and to develop new sustainable growth paths remains very unevenly distributed (Heimeriks et al., 2019). Regarding European policy instruments and regional inequalities, the main target of the cohesion policy is to support economic and social cohesion by reducing disparities between regions and focusing on less developed territories, which receive the largest share of funding. Remarkably, there is an inconsistency between the relatively higher need to promote innovation in these less developed regions and their lower capacity to absorb available funds and successfully invest in innovation activities compared with more advanced regions, or what has been described as the 'innovation paradox' (Gianelle, Guzzo & Mieszkowski, 2020; Oughton, et al., 2002).

Knowledge production is also path and place dependent, where new activities tend to emerge and develop in a region in fields closely related to existing local activities. It is differentiated among locations and every region has its own, unique knowledge base (Heimeriks & Boschma, 2014). There is clear evidence that countries and regions are more likely to diversify into related activities. Heimeriks et al., (2019) showed that the existing scientific portfolio of regions offers opportunities for related diversification and discourages the creation of knowledge on topics unrelated to the local knowledge base. Asheim, Boschma & Cooke (2011) use the term *related variety*, referring to shared and complementary knowledge bases and competences. This concept most probably occurs through knowledge transfer mechanisms such as firm diversification, spinoff activity, labour mobility and social networking. It links knowledge spillovers to economic renewal, new growth paths and regional growth and, if pervasive, it implies that the long-term development of regions depends on their ability to diversify into new applications and new sectors while building on their current knowledge base and competences.

The Quadruple Helix model (QH) constitutes a central element in the design of smart specialization strategies. It promotes the exchange of knowledge creation by bringing together companies, universities or research centres, civil society, independent inventors, and lead users to strengthen the regional innovation system (Carayannis and Grigoroudis 2016). The model forms an integral part of European innovation policy, which aims to create sustainable and inclusive growth in Europe. It situates the role of civil society and citizens as especially valuable for the establishment of social innovations in regions (Carayannis and Campbell 2009). Despite the strong emphasis on the QH model, it is still far from a well-established concept in innovation research and policy, and civil society participation in RIS3 has remained low (Roman et al., 2020).

The major mechanism for bringing actors together in RRI policy is public engagement, one of the European Commission's (EC) six RRI 'keys' along with ethics, gender equality, governance, open science, and science education. The EC describes its RRI policy as a diverse set of societal actors that "*work together during the whole research and innovation process to better align both the process and its outcomes with the values, needs, and expectations of society*" (European Commission, 2018).

Previous initiatives establishing the RRI concept into RIS3 policy making including the MARIE[1] project (Mainstreaming Responsible Innovation in European, S3) pursued the creation of greater awareness among regional stakeholders and the wider public on the potential of S3 policies to promote responsible growth. Additional attempts to integrate regional RRI and RIS3 approaches into a responsible and regionally embedded innovation policy has been done by Fitjar, Benneworth, & Asheim (2019). The authors emphasize the complementarities between both approaches, but RIS3 policy is primarily oriented towards regional competitiveness and therefore does not fully incorporate local institutions and notions of social value, needs or choice – the main concerns of RRI. Conversely, RRI theory, policy and practice does not pay attention to the spatial dimension of innovation processes, which is central in RIS3 approaches. In that sense, RRI ignores the various ways in which the regional context affects not only the development of innovation but also the perception of what is responsible and socially desirable, understanding that knowledge and resources which are necessary for innovation – labour, mobility, R&D collaboration – are to a large extent regional. The lack of social focus in the RIS3 has been also addressed from the social innovation (SI) perspective (Nogueira, Pinto & Sampaio, 2018; Spiesberger, Seigneur & Gómez Prieto, 2018). RIS3 and SI are both largely policy-directed and practice-directed concepts which are instrumentally constructed, in which also actors not traditionally associated with innovation (public service organisations, users, citizens, individuals and social enterprises) can contribute (Richardson, Healy, & Morgan, 2014). There is a social side in smart specialisation that seeks the engagement, inclusion, and empowerment of individuals, while it promotes regional specialisation and development. Citizens and user groups should be considered as important players, both for the identification of social needs and for development and testing of new solutions (European Commission, 2014).

In the following sections, we introduce a novel approach to support the creation of more open, inclusive, and self-sustaining R&I ecosystems in the healthcare and innovation sector. The approach combines insights from RRI with research and innovation smart specialisation strategies. The exercise entails identification of healthcare and innovation stakeholders, the characterisation of the policy landscape in each territory, and the analysis of regional capabilities (strengths) that provide opportunities for innovation processes. Additional recognition of regional needs allows us to assess the alignment of the selected needs with respect to the regions' capabilities and current policy mix.

METHODOLOGY

By using a mixed method strategy that combines qualitative and quantitative analyses, our approach examines three different dimensions: *stakeholders, policies, and R&I strengths*. Through the identification of **local stakeholders** by local partners involved in the CHERRIES project a network of actors was built acknowledging the 4P model of interest conformed by providers, practitioners, payors, and policymakers (Ritz et al., 2014), and further enriched using the quadruple helix of innovation. The regional consortium employed local criteria to select key stakeholders and defined their roles in the project. In parallel, they further specified the principal national and regional **policy frameworks** on RRI, healthcare, and science and innovation. Afterwards, the analysis of the **knowledge and innovation base** used the RIS3 as a reference for the identification of regional priorities by showing the scientific fields or areas where each region has a higher level of specialisation and could therefore be used as a driver for the innovation process. The knowledge

1 https://www.interregeurope.eu/marie/

base covered scientific articles, registered patents, and European projects across all disciplines, but particularly in the biomedical and health science field as a representation of the healthcare sector. The present paper only refers to the analysis of scientific articles, and uses bibliometric indicators based on CWTS internal database (Web of Science's (WoS) produced by Clarivate Analytics). We calculated the Revealed Comparative Advantage (RCA) based on relatedness by analysing regional publications (2014-2018) as an indicator of the scientific fields or areas in which the region has an above-average concentration of publications compared to other European regions (Hidalgo et al. 2007). Furthermore, we identified which scientific fields are often found together in the same region, as a representation of the ability of the territory to diversify into related areas of expertise. Complementarily, the employment of the micro-fields level analysis provided a more detailed characterization of each prioritised field by providing information about scientific disciplines, relevant topics, and even specific diseases or disorders. The micro-level analysis method uses an algorithm, where each publication is assigned to one of the 4,013 fields based on a large-scale analysis of hundreds of millions of citation relations between publications. These micro-level fields are embedded into the five main fields of science, namely: social science and humanities, mathematics and computer science, biomedical and health science, physical science and engineering. For further methodological details please refer to Waltman and Van Eck (2012). The characterisation of the selected fields also considered the most representative journals in which the region publishes, together with the publication content by using the titles of articles contained in each micro-level field. For those with a larger set of publications we used text mining techniques or term maps (Vosviewer[2]) to detect the core topics in the abstracts.

RESULTS AND DISCUSSION

STAKEHOLDER NETWORK AND POLICY MAPPING

In the context of CHERRIES project implementation and as a result of the regional analyses, the identification of stakeholders for the Murcia region found 84 institutions. These actors are mostly represented by civil society organisations (CSO) linked to patients' associations and hospitals. Cyprus identified 50 actors with hospitals and health centres (providers) and higher education institutions. For Örebro, 58 actors were reported, most of them belonging to public administration organisations (policymakers) and CSOs. Stakeholders from the private sector composed of firms, start-ups, and SMEs or payors were the least represented in the three regional networks. In this regard, the regional partners underlined some difficulties arising from stakeholders' identification and engagement process from the private sector, which could also suggest that the business and innovation system is detached from the regional (scientific) knowledge production and from the public sector. Additionally, the development of similar previous European projects in the field of health and innovation in the region facilitated the stakeholder mobilisation process. This was for instance the case of Murcia region and the InDemand project[3].

With respect to the compliance of the quadruple helix of innovation, the broad involvement of society organisations characterised predominantly by patient associations in the case of Murcia and Örebro highlights the relevance of co-creation processes where the citizen/end-user perspective is integrated into the innovation cycle. This core principle represents an essential focus of the RRI perspective. On the contrary, the 4P model of healthcare proved to be insufficient in portraying the diversity on the institutional landscape in the regions. It disregards the essential role of academia (universities and research centres) in knowledge dissemination and its contribution to innovation dynamics.

Based on the results of the policy mapping exercise, there is no overarching RRI-policy in the regions. For the three territories, the most developed RRI keys were *Gender equality*, and *Open Science* converging towards Open access, with national or regional policies in place. Örebro was the region with the most diverse RRI keys among policy frameworks, with science literacy and scientific education (SLSE) and ethics keys covered in their policy instruments. As such, RRI does not appear to be grounded as a concept in the territories, however, RRI practices can be easily found. Concerning the ethics RRI key, bioethics is generally regarded as biomedical and clinical research, yet the more general concept of integrity is not addressed at the policy level (except in Örebro). The Swedish research strategy has three overarching guidewords to indicate the future choices: 'Knowledge, quality and integrity', also including strengthening and coordination of science communication, and new infrastructures for knowledge dissemination (SLSE RRI key). From a policy mix perspective, in Cyprus and Spain no reference to science literacy and science education could be found. Likewise, in all three regions there is no specific mentioning on how research should engage with the public stakeholders (public engagement). In the current situation, European policy has translated this lack of public engagement and communication into the 'new' citizen science policy perspective. This raises questions since the traditional citizen science is about citizens supporting science initiatives, albeit the European idea refers to the public having access to and engaging in science, in a less 'data collectors' manner. Only Örebro region is mentioning citizen science as a new policy avenue. In this regard and during a reflection session carried out in 2020 (interregional workshop) addressing RRI needs and potentials, the common issue stressed within the three regional focus groups was the necessity to create permanent space or "arenas" for dialogue and deliberation. A collaborative space is essential for societal engagement in order to make decisions regarding the way healthcare services are provided, which technologies are developed and adopted, and how services are organised. This space is particularly relevant during the needed identification process and definition of innovation priorities in health (and in general in the territory) in a more open, inclusive, responsive, and socially aligned manner. This aspect can be considered as a first important indication for institutional RRI changes to promote at the regional level, with the active collaboration of key actors from the different innovation communities.

2 Vosviewer is a software for constructing and visualizing bibliometric networks: https://www.vosviewer.com/
3 https://www.indemandhealth.eu/indemand-murcia/

IDENTIFICATION OF REGIONAL PRIORITIES

SCIENTOMETRIC ANALYSIS AND REGIONAL STRENGTHS

Figures 1 and 2 illustrate the outcomes for Murcia region, as an example of the results for the scientometric analyses performed in each region intended to support the identification of the regional strengths. This region constitutes an interesting case study due to the clear connection revealed between the topics identified as capabilities and the priorities addressed by the RIS3 instrument. In contrast, the analyses exposed a misalignment between the policy agenda and the issues that surfaced from the Call for Needs-process aiming to recognise local needs in healthcare and innovation. A visualization of the prioritized fields is shown in Figure 1, which corresponds to the measurement of the revealed comparative advantage (RCA). It encompasses the years 2014 to 2018 and analyses a total of 14,433 publications including articles, reviews, and conference proceedings.

In the case of Murcia, fields in relation to environmental science and agriculture shape a dense area in Figure 1 (left side of the image - green cluster). Some of the relevant fields concerning the biomedical and health science field are food science and technology, ophthalmology, dentistry, oral surgery and medicine, cardiac and cardiovascular system, urology and nephrology, hematology, immunology (at the bottom of the image - blue cluster). Highlighted fields at the interface of health and social science are rehabilitation, sport science, nursing and psychology/psychoanalysis (right side of the image - red cluster). Additionally, further information to be extracted from Figure 1 concerns the proximity of the fields (nodes) in which the region would have a better chance to specialize based on its current skills. As an example, as Murcia performs well in Rehabilitation (marked with a black circle), the region could diversify their skills into closer and related fields such as sport science, that appear adjacent to this field in the image below.

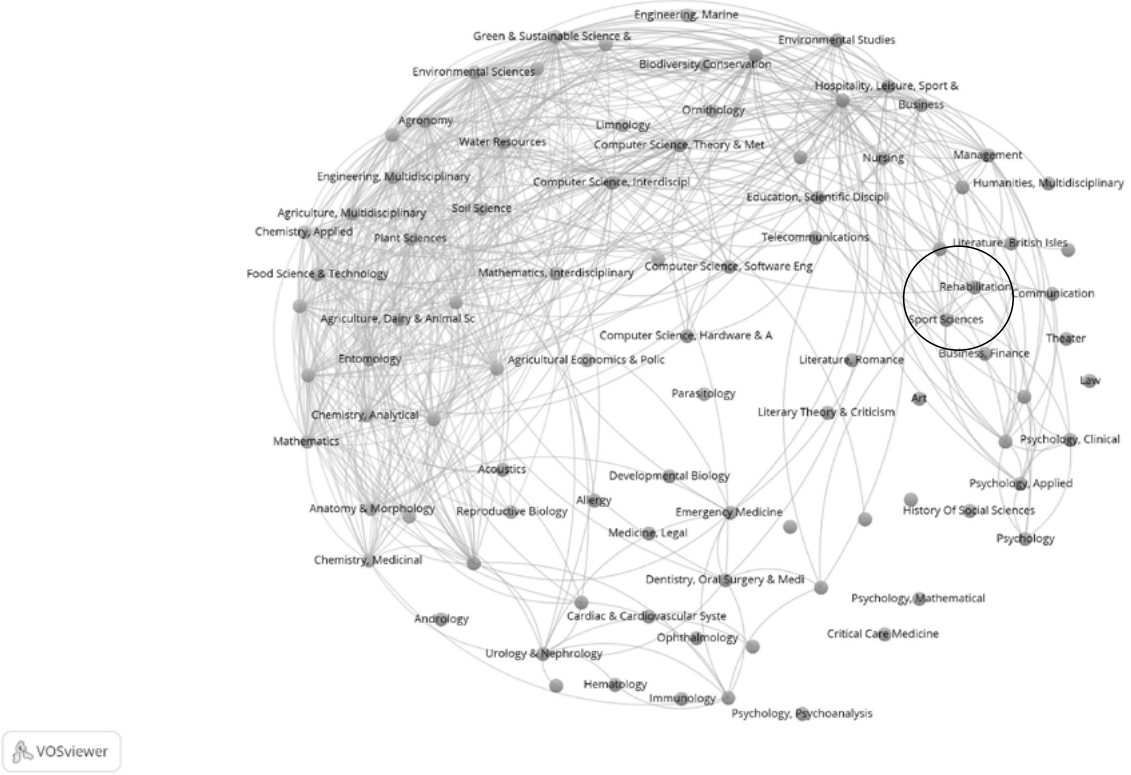
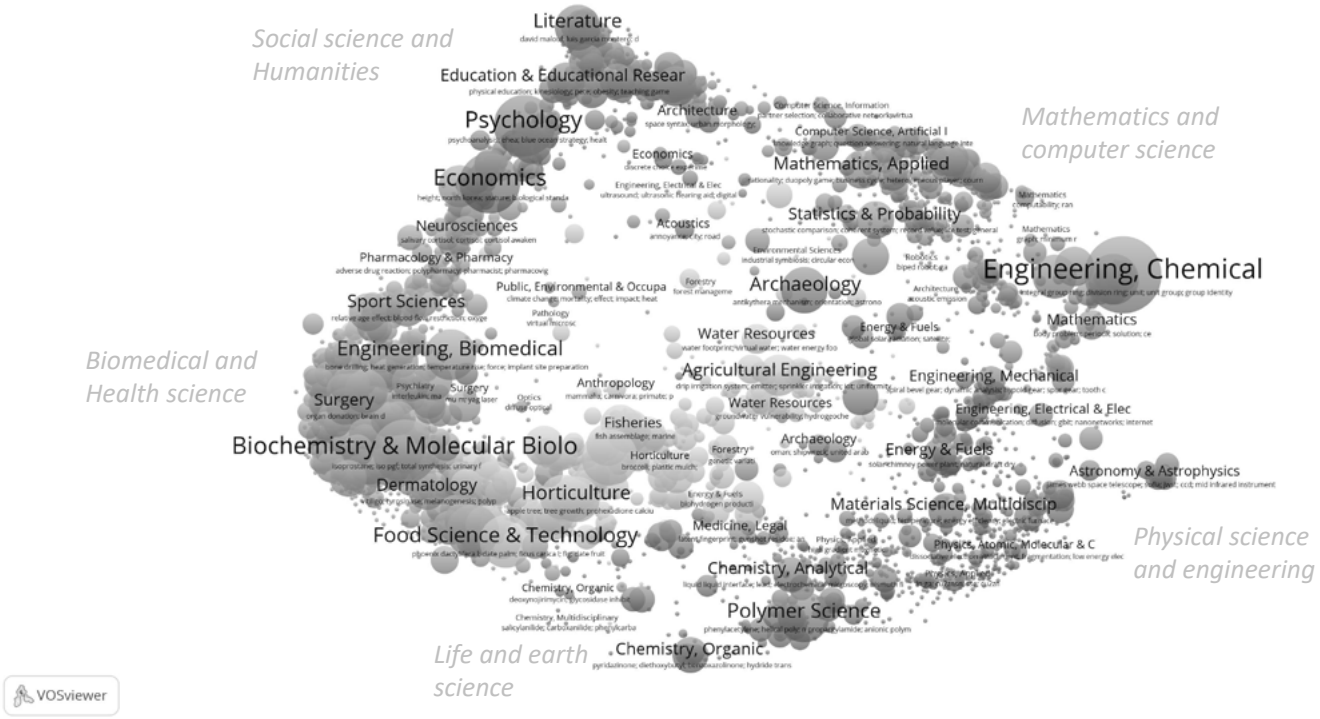


Figure 1. Results of relatedness analysis depicting fields of science with a Revealed Comparative Advantage (RCA) in Murcia.

An additional example of the scientometric analyses performed is the following visualisation portraying the most relevant scientific fields and topics in Murcia. The colours represent the main fields of science, and each circle symbolizes a micro-level field, where the bigger the circle (node), the higher the number of publications produced in that specific micro-field. Figure 2 presents an overview of the topics arising as significant from each main field (e.g., social science, engineering, health science). As stated in the previous smart specialization report, Murcia region has an extensive development of the fields related to agriculture, plant science and nutrition, and environmental science. Similarly, the micro-level field analysis shows a high relative number of publications or level of specialization on these fields and also in connection to Biochemistry & Molecular Biology and Pharmacology & Pharmacy fields.

Figure 2. Results of the micro-level scientific field analysis using the relative number of publications for Murcia.



ALIGNMENT OF REGIONAL POLICIES, STRENGTHS AND NEEDS

Identifying local needs in the context of a European project shows that strengths in one particular area using the knowledge base do not necessarily relate or align to local healthcare practices and innovative activities. The contribution of local capabilities and their active use require adaptation and adoption to local cultures. To balance misalignment and create synergy at the local level is not happening per se and adjustment of stakeholders' dynamics requires active engagement. In the same way, local ownership and commitment develop gradually over time and seldom refer to RRI or innovation, but rather to specific activities. The following table summarises the results obtained from the policy mapping exercise, the identification of the capabilities (strengths) employing scientometric analyses, and the priorities and needs defined at the regional level by the "Call for Needs" process.

Table 1. Overview of the policy topics for RIS3 and RRI policies, the strengths, priorities and needs identified for each region.

REGION/COUNTRY	POLICY		STRENGHTS	NEEDS
	RIS3	RRI policy	Knowledge base analysis	Call for Needs
Murcia - Spain	Agriculture, plant science and nutrition, environmental science. quality of life for well-being	Gender equality, open access, science literacy and scientific education (SLSE)	Environmental science (agri-food chain): dietetics and nutrition. Biomedical: ophthalmology, dentistry, oral surgery, sport science	Early detection of progression in multiple sclerosis
Cyprus	Health, ICT and biomedical applications, e.g., early warning, diagnosis, and early medical care provision.	Gender equality, open access	Genetics and heredity pediatrics, rehabilitation, biomedical social science, psychology, cardiology, nursing	Provision of medical services to citizens living in rural and remote areas
Örebro (Sweden)	Health and social care, open social efforts, accommodative health care. Health robotics	Gender equality, open science, science literacy and scientific education (SLSE) and ethics	Gerontology, nursing, psychology/ psychiatry, gastroenterology and hepatology, automation & control (robotics)	Involuntary loneliness among elderly

MURCIA

For Murcia we could observe that innovation policy instruments such as S3 established clear leadership in sectors such as the agri-food value chain, including agriculture, livestock, fishing and the food industry, and the environmental field as well. Organisations are primarily economically specialized in agricultural inputs and services. Likewise, the analysis of the knowledge base (RCA and micro-level fields) for the region showed similar topics as strengths for Murcia. In its S3 document, Murcia has defined broad health-related priorities with a focus on the quality of life and wellbeing.

As a result of the call for needs-process in Murcia, the region received proposals for the treatment of different chronic illness (e.g., lumbar and cervical pathology, osteoarthritis, pelvic floor disorders), and after the selection process, early detection of progression in multiple sclerosis was given priority. It is worth noting that the call for needs-process did not specify a particular topic to receive the local demands. Instead, it was open to reveal potential new demand-driven needs for health-related innovations in the territory. This may have had a direct influence on the detachment of the needs detected by the region and the strengths identified from the knowledge base and the RIS3 instrument for Murcia region. In this way, the potential strengths detected in Murcia do not overlap with the expressed priorities and needs, albeit chronic illness could be treated by health promotion such as food and nutrition and sports. It should be emphasized that Spain has the highest life expectancy in the EU and social inequalities in health are less pronounced than in many other countries. However, many years of life in old age are lived with some chronic diseases and disabilities, increasing demands on health and long-term care systems.

CYPRUS

The regional priorities of Cyprus are partially aligned with the needs detected. It was possible to observe (Table 1) that topics addressed by RIS3 policy are in line with the priorities and needs identified by the region. Cyprus has defined health concerning ICT and biomedical applications as a priority in their S3 document. As a result of the feedback obtained during the stakeholders' engagement process and demands identified by Cyprus in the call for needs, the selection indicated telemedicine as a local demand with a special focus on the provision of medical services to citizens living in remote areas who do not have easy access to healthcare services and prescribed medicines. The topic matched with the areas addressed by RIS3 policy framework (ICT and biomedical applications), however, did not employ directly the existing capabilities in the health science fields (nephrology, cardiology, paediatrics, rehabilitation, psychology) – fields for which Cyprus was not aware of its potential. In this context it is relevant to note that in healthcare the public sector is dominant. The links between the public sector and the R&I system are less developed, and therefore smart specialised, RRI-based innovation develops less easily.

ÖREBRO

The region has defined health and social care in its S3 priorities strategy. The biomedical and health science field analysed by the relatedness analysis (RCA) from their knowledge base supports the health innovation area. Within this area, the fields of expertise align well with social care such as gerontology, nursing, psychology/psychiatry. Additionally, health robotics as a priority could build on a sophisticated knowledge base within the automation and control field. The topics covered by the regionally submitted needs addressed demands for social contacts among the elderly to tackle loneliness, together with the development of technical skills to use digital tools to counteract this issue. Thus, the demand anticipated by the project partners resulted in bottom-up, demand-driven health needs, which align well with their territorial strengths and broader priorities. In Örebro, the responsibility for the provision of health care to the elderly is shared by the county's municipalities and region Örebro county. The recent "Swedish Government Official Reports" (SOU) report (2020) also refers particularly to elderly care during the pandemic and reflected on the un-preparedness of the health system.

It is worth mentioning that one challenge remains in this territory: the actors that could bring together knowledge, innovation and a healthcare miss possibilities for public engagement among the regional stakeholders. This aspect has been emphasized not only for Örebro region, but for Cyprus and Murcia as well.

RRI IMPLEMENTATION

Awareness of RRI varies considerably across stakeholders, many having no prior knowledge of the concept. However, the overall impression is that there is a positive attitude towards orienting territorial research and innovation systems in RRI terms, however, we identified substantial differences across stakeholders regarding how RRI can be framed. Commonly, stakeholders frame RRI intuitively, from their personal experiences and world views (e.g., in terms of research integrity), or align it to a dominant discourse within their organizations (e.g., CSR for the industry). Yet, common key elements of a perceived RRI approach emerge. Many expressed the view that scientific research and innovation should be oriented towards societal needs in the region and be connected to society with territorial actors. Stakeholders often stress principles of inclusion, deliberation and reflection through collaboration and participation (e.g., co-creation) and continuous, open dialogue between different actors and society.

CONCLUSION

Our approach can assist policymakers and other actors in designing and implementing RIS3 strategies that respond to local needs and preferences. By combining information on the relative strength of regional knowledge production activities (e.g., science and technology outputs) with information about regional stakeholders, local needs, and policies, we can specify priorities that can help to maximise the regional development potentials. Furthermore, our analyses show that scientific capacities that could be useful for regional development do not necessarily align with the demand-driven regional needs. Demand driven research priority setting for funding schemes is very much in sync with RRI, nevertheless, the smart specialisation paradigm does not always adequately include regional needs. Notable is that actors formulating demand-driven needs are not always aware of potentially interesting local knowledge for innovative developments. The first and foremost step in supporting territorial RRI is engagement and understanding local cultures. Recommendations

should move towards searching for innovation opportunities on the basis of RRI-based local needs, in view of the local strengths of the knowledge base. It is about the translation of project contexts to local contexts and making sure that changes remain when projects are gone.

The integration of qualitative and quantitative methods to understand territorial specific characteristics constitutes a novel and promising approach, conveying regional relevant scientific and technological information that was previously unavailable, and link it to the regional priorities. The overall application of this approach appears highly beneficial with still some opportunities for enhancement.

Further efforts aimed at the integration between the RIS3 and RRI policy approaches are necessary for a better social alignment of the innovation decision-making process by establishing bridges between existing and new territorial actors of the regional R&I healthcare system. This includes in the process different knowledge perspectives and creates the conditions for the building of collective responsibility toward responsible innovation in health or "territorial RRI".

REFERENCES

Asheim, B., Boschma, R. & Cooke, P. (2011) Constructing Regional Advantage: Platform Policies Based on Related Variety and Differentiated Knowledge Bases, Regional Studies, 45:7, 893-904, DOI: 10.1080/00343404.2010.543126

Boschma, R., Heimeriks, G., & Balland, P. (2014) Scientific knowledge dynamics and relatedness in biotech cities. Research Policy, 43, 107–114. DOI:10.1016/j.respol.2013.07.009

Carayannis, E. & Campbell, D. (2009) Mode 3' and 'Quadruple Helix': Toward a 21st century fractal innovation ecosystem. International Journal of Technology Management 46: 201. DOI: 10.1504/IJTM.2009.023374

Carayannis, E., & Grigoroudis, E. (2016) Quadruple Innovation Helix and Smart Specialization: Knowledge Production and National Competitiveness. Foresight and STI Governance 10: 31–42. DOI:10.17323/1995-459x.2016.1.31.42

Dahl Fitjar, R., Benneworth, P. & Terje Asheim, B. (2019) Towards regional responsible research and innovation? Integrating RRI and RIS3 in European innovation policy, Science and Public Policy, Oxford University Press, vol. 46(5), pages 772-783. DOI: 10.1093/scipol/scz029

European Commission (2012) Guide to Research and Innovation Strategies for Smart Specialisations (RIS 3) (document)

European Commission (2014) National and regional innovation strategies for smart specialisation (RIS3) - Cohesion Policy 2014-2020 (document)

European Commission (2018) Horizon 2020 - Work Programme 2018-2020 - Science with and for Society (document)

Spiesberger, M., Seigneur, I., & Gómez Prieto, J. (2018) Smart specialisation and social innovation : from policy relations to opportunities and challenges : evidence from six case studies on clean energy regional initiatives, Publications Office. European Commission, Joint Research Centre. DOI: 10.2760/601340

Foray, D., David, P. & Hall, B. (2011) Smart specialisation from academic idea to political instrument, the surprising career of a concept and the difficulties involved in its implementation (paper)

Foray, D. (2016) On the policy space of smart specialization strategies. European Planning Studies, 24(8), 1428-1437. DOI: 10.1080/09654313.2016.1176126

Gianelle, C., Guzzo, F. and Mieszkowski, K. (2020) Smart specialisation: what gets lost in translation from concept to practice, Regional studies, ISSN 0034-3404, 54 (10), p. 1377-1388, JRC116340. DOI: 10.1080/00343404.2019.1607970

Heimeriks, G., & Boschma, R. (2014) The path- and place-dependent nature of scientific knowledge production in biotech 1986–2008. Journal of Economic Growth, 14, 339–364. doi:10.1093/jeg/lbs052

Heimeriks, G., Deyu, Li, Wout, L., Meijer, I. & Yegros, A. (2019) Scientific knowledge production in European regions: patterns of growth, diversity and complexity. European Planning Stud. 27(11):1-21. DOI: 10.1080/09654313.2019.1645814

Hidalgo, C., Klinger, B., Barabási, A. & Hausmann, R. (2007) The product space conditions the development of nations. Science, 317(5837), 482-487. DOI: 10.1126/science.1144581

Nogueira, C., Pinto, H. and Sampaio, F. (2018) Social innovation and Smart Specialisation: Challenges and Opportunities for Atlantic Regions. European Public & Social Innovation Review. 2, 2 (May 2018), 42-56. DOI: 10.31637/epsir.17-2.4

Nomaler, Ö, Frenken, K., & Heimeriks, G. (2014) On scaling of scientific knowledge production in U.S. metropolitan areas. PLoS One, 9, e110805. DOI:10.1371/journal.pone.0110805

Richardson, R., Healy, A., & Morgan, K. (2014) Smart Specialisation for Regional Innovation: WP2 Embracing Social Innovation. Cooperation Programme: Social Sciences and the Humanities

Ritz, D., Althauser, C., Wilson, K. (2014) Connecting Health Information Systems for Better Health: Leveraging interoperability standards to link patient, provider, payor, and policymaker data. Seattle, WA: PATH and Joint Learning Network for Universal Health Coverage (Document).

Roman, M., Varga, H., Cvijanovic, V., & Reid, A. (2020) Quadruple Helix Models for Sustainable Regional Innovation: Engaging and Facilitating Civil Society Participation. Economies, 8(2), 48. DOI:10.3390/economies8020048

OECD (2013) Innovation-driven Growth in Regions: The Role of Smart Specialisation, Paris, OECD Publishing (Document)

Oughton, C., Landabaso, M., & Morgan, K.J. (2002) The Regional Innovation Paradox: Innovation Policy and Industrial Policy. The Journal of Technology Transfer, 27, 97-110. DOI: 10.1023/A:1013104805703

Waltman, L., & Van Eck, N.J. (2012) A new methodology for constructing a publication-level classification system of science. Journal of the American Society for Information Science and Technology, 63(12), 2378–2392. DOI:10.1002/ASI.22748

AUTHORS

SONIA DANIELA MENA JARA
Leiden University. Centre for Science and Technology Studies
Kolffpad 1, 2333 BN Leiden
E: s.d.mena.jara@cwts.leidenuniv.nl
ORCID: 0000-0002-2153-2431

INGEBORG MEIJER
CWTS Leiden University
Kolffpad 1, 2333 BN Leiden
E: i.meijer@cwts.leidenuniv.nl
ORCID: 0000-0003-1481-1739

GASTON HEIMERIKS
Copernicus Institute of Sustainable Development, Utrecht University
E: g.j.heimeriks@uu.nl
ORCID: 0000-0002-0577-6938

TIM WILLEMSE
CWTS Leiden University
E: t.p.h.willemse@cwts.leidenuniv.nl
ORCID: 0000-0002-3778-8247

Commentary

EVALUATIVE CONVERSATIONS: TRANSLATING BETWEEN DIVERSE STAKEHOLDERS IN REGIONAL RRI PROJECTS

TJITSKE HOLTROP, INGEBORG MEIJER, PAULA OTERO-HERMIDA, ANESTIS AMANATIDIS, CHIARA BUONGIOVANNI, DONATELLA CASALE, CLAUDIA COLONNELLO, ALESSANDRO DESERTI, FABIO FEUDO, ALAN HARTMAN, MARIANNE HÖRLESBERGER, ILDIKO M. IPOLYI, NHIEN NGUYEN, MIKA NIEMINEN, GABRIELE QUINTI, TINE RAVN, FRANCESCA RIZZO, FELICITAS SCHMITTINGER, NORBERT STEINHAUS, ROGER STRAND, THOMAS VÖLKER, RENE WINTJES AND EMAD YAGHMAEI

DOI: 10.22163/fteval.2022.544

ABSTRACT

Since the summer of 2020, researchers from ten projects pertaining to the Horizon2020 Science with and for Society (SwafS) call have been meeting virtually as the SwafS14 Monitoring and Evaluation ecosystem. Topics of discussion were the trials and tribulations of their regional Responsible Research and Innovation (RRI) projects as well as their strategies for monitoring and evaluation. In this paper we make a first attempt at presenting these issues as problems of translation between different kinds of stakeholders. After an exploration of the diversity of stakeholders and the process of translation in regional RRI, we suggest *evaluative conversations* as a way of improving regional RRI. We intend to develop this idea in the future and that these conversations will facilitate more responsible and engaged monitoring and evaluation and contribute to better R&I policies.

INTRODUCTION

What is the best way to conduct evaluations of regional Responsible Research and Innovation (RRI)? We, the authors of this article, have been discussing this question since the summer of 2020 in a series of conversations. We represent 11 projects, all funded by the EU Horizon2020 Science with and for Society (SwafS) program. Ten projects – SeeRRI, TeRRItoria, TeRRIfica, SISCODE, CHERRIES, DigiTeRRI, RRI2SCALE, TRANSFORM, TetRRIs, and RIPEET –responded to the EU Horizon 2020 SwafS14 call "Supporting the development of territorial Responsible Research and Innovation" with the strategic aim to foster RRI in regional and local science and innovation systems. The 11th project – SUPER-MoRRI – is the host of our conversations. SUPER-MoRRI focuses on monitoring and evaluating RRI. It intends to develop an evaluation framework that adapts metrics to their specific contexts and actors, thereby promoting responsible use and interpretation of the results.

RRI is defined by the European Commission as the ambition to let diverse groups of societal actors (researchers, citizens, policymakers, entrepreneurs, social innovators, third sector organizations, etc) "work together during the whole research and innovation process in order to better align both the process and its outcomes with the values, needs and expectations of society." Our projects address regional RRI ("territorial RRI" in EC jargon), which refers to the support of more open, inclusive, responsive, and reflexive regional and local science and innovation systems to improve the governance of regional transformations and the response to regional challenges. Our projects generally aim to align science, innovation, and society by facilitating collaboration between local stakeholders such as research performing organizations (RPO), higher education institutions (HEI), public authorities, civil society organizations (CSO), and, to a lesser extent, research funding organizations (RFO), small and medium sized enterprises (SME), and industry. These stakeholders represent around 35 regions from 20 countries across Europe.

We came together as a group to discuss the possibility of a shared M&E plan, cocreated with the regional partners. Background for this were the following assumptions: Collaboration between regional RRI projects would provide detailed information on the monitoring of regional research and innovation projects with respect to the RRI keys developed in the MoRRI project and mentioned in the original EU Horizon 2020 call: Gender Equality, Science Literacy, Public Engagement, Ethics, Open Access, and Governance (Technopolis, 2020). Other "indicators" besides these keys might be added as well, possibly reflecting conditions such as sustainability and the ARRI process dimensions (anticipation, reflection, responsiveness, inclusion. Stilgoe, Owen, Macnaghten, 2013). Also, Sustainable Development Goals (SDG) and Smart Specialization (S3)-related indicators were expected to be part of the process.

When our meetings started in the summer of 2020 some of the projects had only just started, while others had been running for some time. All projects were struggling with the effects of COVID19. We soon realized that the differences between the approaches, emphases, collaborations, or timelines of the different projects were too big to consider a one-size M&E plan to fit all projects. We collectively decided to explore our differences further to see what other output or activities could follow. The SUPER MoRRI team compared several existing M&E plans (more about this later), from which four general differences emerged: 1) the RRI frameworks that projects draw on, 2) the diversity of stakeholders and how to engage them, 3) the evaluation practices adopted by the projects

and, 4) the indicators, what they mean and how to use them. We used the material collected in these conversations as a starting point for the analysis of RRI projects and their evaluations as processes of translations. This analysis inspired us to suggest "evaluative conversations" as answer to some of the translational issues in (evaluating) RRI.

In the following, we will first discuss the conversations we had around the general differences of RRI framing, stakeholders and how to engage them, evaluation practices and indicators. Secondly, we will reflect on these issues as problems related to translation, which reveal two crucial elements: 1) the diversity of stakeholders, and 2) a process of subsequent steps to integrate this diversity successfully and effectively. We will, thirdly, explore these two elements drawing on insights and building blocks from other projects and initiatives, to finally introduce our idea of evaluative conversations. We emphasize that this paper is a first attempt at understanding SwafS14 projects as projects of translating between heterogeneous stakeholders that could be strengthened by the approach of the evaluative conversations. We hope to elaborate on SwafS projects as projects of translation and on evaluative conversations in the future.

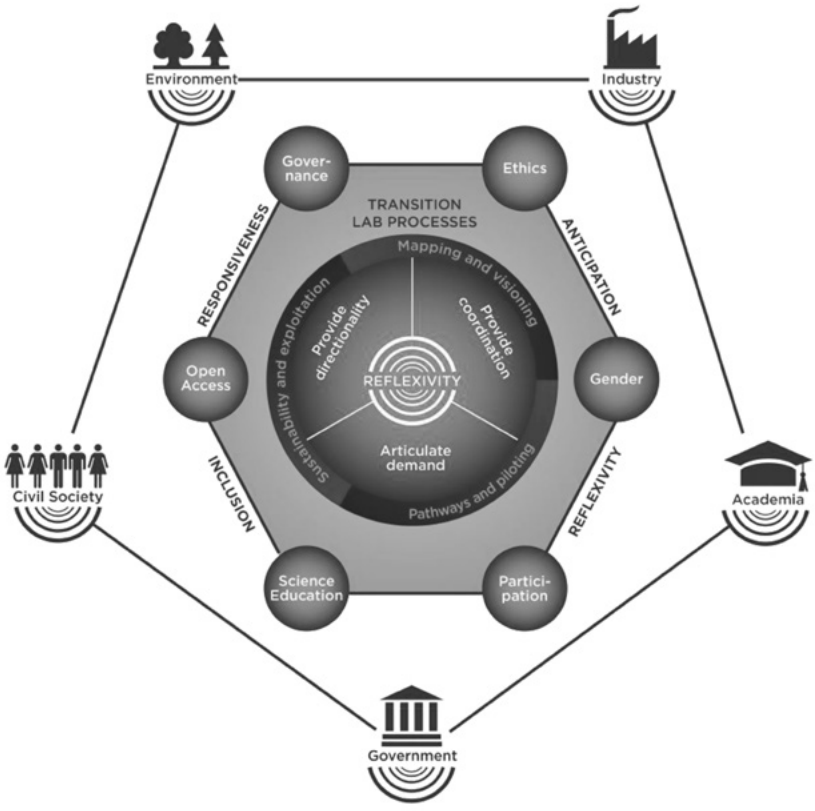
EVALUATING REGIONAL RRI: WHAT ARE THE ISSUES?

In early 2021 we organized four focused conversations around one of the four issues we mentioned above. In the first focus group we invited the participants to reflect on how their own projects related to RRI frameworks and responsibility. The image here, taken from the RIPEET project, represents an RRI constellation based on the MoRRI keys of ethics, gender, participation, science education, open access, and governance; the RRI dimensions of anticipation, reflexivity, inclusion, and responsiveness; and the quintuple helix stakeholders of industry, academia, government, civil society, and environment. In particular, the following two questions were discussed: How to incorporate all these skills, elements, and stakeholders successfully in a project? And what kinds of RRI-inspired trajectories of change are being pursued?

From the conversation around RRI frames we became aware of the diversity of approaches to RRI that our projects work with. Within each project a different focus, selection of stakeholders, ambition, and/or territory can be observed, with different aims as well. Some projects have a strong focus on systemic institutional or organizational change. Others focused on citizen engagement as crucial condition for grounding RRI. Again others argued that certain keys, such as governance, permeated all aspects of regional RRI. There was no consensus on the quintuple helix where some argued that the environment cannot be considered an actor. Instead, we should stick with the idea of the quadruple helix and consider sustainability to be an overarching feature of RRI. Participating researchers from all projects shared the opinion that response-able and inclusive engagement with a host of relevant, yet heterogeneous actors, concerns and languages is both a key feature and challenge of RRI.

Amanatidis and Meijer offer a useful framework for making sense of the diversity (SUPER MoRRI, 2021). They conducted a survey among 29 SwafS funded RRI projects. They analyzed the diversity they encountered in terms of three kinds of RRI. They refer to these kinds of RRIs as multistabilities: relative stable results of processes in which many actors, ambitions, resources, and contexts come together around one of many possible versions of RRI. The first (1) are projects that create value *for* specific societal groups. RRI projects offer resources such as knowledge, funding, and networks, encourage stakeholders to define their preferred

Figure 1: RRI constellation based on RRI dimensions, MoRRI keys and quintuple helix stakeholders. http://ripeet.eu



transformations, and facilitate the process towards this direction. In such constellations, the challenge is to connect ideas and practices of RRI to the worlds of local stakeholders. The second (2) kind of RRI project are those that focus on democratizing research and innovation, by creating common views across all actors and holding each actor equally accountable for decisions and actions taken throughout the process. For these projects it is a challenge to balance the expectations of non-consortium members, especially regarding the time and effort the engagement takes. The third (3) kind of RRI project is concerned with mobilizing actors around already existing RRI conceptualizations. In these projects, a lot of investment is usually needed to make the RRI framework accessible to stakeholders.

The second conversation was about the engagement of stakeholders as key concern and *strategy* for RRI. Engaging stakeholders effectively is not straightforward. It is important to keep in mind that these regional RRI projects work with different kinds of stakeholders. There are often multiple regional and non-regional partners involved in the project consortium who organize and manage regional coalition building. Then, there are regional stakeholders who become part of these coalitions, while having their own stakeholders as well. So there was a need to discuss how to involve all these actors, what is at stake for them and who makes the decisions.

Finding and enrolling stakeholders is a complicated process. Keeping them committed for a longer period even more so. Issues are the building of trust between stakeholders, understanding each other's concerns and ambitions within the language of RRI, subscribing to the need and method of regular assessments, and all in the context of additional COVID19 restrictions. The drive to contribute to change and innovation ideally comes from the stakeholders, but this is not always the case. Commitment needs a lot of maintenance and care, communicating the benefits of RRI and projects protocols to stakeholders and translating between stakeholders' different needs and languages.

The third conversation was about evaluation logics and practices. If conducting regional RRI projects is already complicated, then what about evaluating them? Not all projects had designed a monitoring and evaluation plan yet, or at all. We compared four available evaluation plans – of SISCODE, CHERRIES, TeRRItoria, and SeeRRI – for their purposes and justification of M&E; the approaches to M&E; and the tools and instruments applied. Table 1 below gives a comprehensive overview of the aims, approaches, and tools.

Table 1 Listing the different elements of our SwafS14 M&E plans

M&E plans	
purposes and justification	long-term durability and sustainability; framework for self-sustaining RRI ecosystems; recommendations on policy and governance structures; feedback to internal, organizational and institutional contexts; demonstration of benefits of RRI; raising awareness of potential challenges; sustainability and transferability; accountability
approaches to and aspects of M&E	problem-solving orientation; developmental evaluation; a quality orientation (evaluation is understood as a managerial procedure); a realist evaluation and co-production model; a formative process evaluation (formative evaluation, inspired by the deliberative democratic evaluation perspective); a summative impact assessment ("theory-based evaluation"); a relational approach to evaluation of social innovation; theories of change (Schwandt, 2015); theories of organizational change
tools and instruments applied	collect evaluative data and input through deliverables and virtual talks; evaluation questionnaires, time series; relevance/ effectiveness/ efficiency/ impact; focus on context-mechanism-outcome; implement a participatory and inclusive approach that relies on partners and stakeholders; attendants' satisfaction on involvement, degree of influence, decision making, transparency of processes, incentive mechanisms, voluntariness, implementation and perceived benefits; inspirational catalogue of indicators for the co-creation of context-specific success criteria; selection of indicators includes relevant MoRRI-indicators, Sustainable Development Goals; on-site visits; discussion/focus groups; semi-structured interviews.

Three themes permeated the conversation about evaluation practices. In line with the other conversations, the problem of engaging stakeholders in evaluation practices existed for many. It remained a struggle to explain the benefits to stakeholders of RRI and the need for regular assessments. A second theme was the difficulty of accounting for regional differences. Many regional RRI projects weave together various interventions or forms of collective experimentation in different regions. Comparing these is difficult, and so is making quantitative statements about the changes seemingly caused by these interventions. Lastly, evaluations have several formal and informal goals and effects. Stakeholders may have different needs or opinions about these. Evaluation protocols are crucial for accountability and governance purposes, but informally they are also a way of being in touch with stakeholders, or a way of raising awareness around RRI, and a means of learning about issues that are occurring in the project and solving them. Sometimes things take time to come to fruition, and at other times good things happen that were not anticipated or are not easily measurable. These phenomena escape the protocolized, formalized style of evaluation that comes in long surveys with closed-ended questions.

The problem with indicators – the subject of the last conversation – is similar to the one we mentioned in the context of evaluation practices. Indicators play different roles in different stages of the RRI and evaluation process. During the project they can point out to partners what might be important or needs to be addressed. After the project, indicators can be used for evaluating the project or communicating project outcomes. Moreover, they are useful in terms of accountability and the conceptualization of new projects and funding applications. Even if indicators are not informative or useful during the project, stakeholders and project partners might still need them to account for and communicate results. A case in point have been the MoRRI indicators, as indicators derived from the MoRRI project have been included in many of the SwafS calls. Since the MoRRI indicators are oriented towards the national level, several projects struggled to use them in their regional contexts. In addition, not all projects focus on the RRI keys and consequently MoRRI indicators are not suitable for those projects. A last issue that needs mentioning, is that institutional or systemic change needs much more time than the duration of our projects. With the available indicators, these benefits of RRI, which take a long time to manifest themselves, can neither be assessed nor predicted.

Two topics stand out in these conversations. The first is the issue of productive engagement of stakeholders in which the translation of the benefits of RRI and the need for regular evaluations play a central role. The second is the confusion around different informal and formal uses and needs of assessments and indicators. In the following, we draw on the sociology of translation to make sense of these issues.

(EVALUATING) REGIONAL RRI AS PROCESSES OF TRANSLATION

The challenge of translation is key to regional RRI and its monitoring and evaluation. In this context, the following questions arise in particular: How to move from RRI conceptual frameworks to innovation done responsibly? How do EU government policies transform regional activities? How to translate effects and change into indicators? How to translate between the needs and wishes of the European Commission, expert networks, regional projects, local partners, and stakeholders? In this section we turn to the sociological concept of translation to understand issues around the engagement of stakeholders and uses of indicators and evaluation systems better.

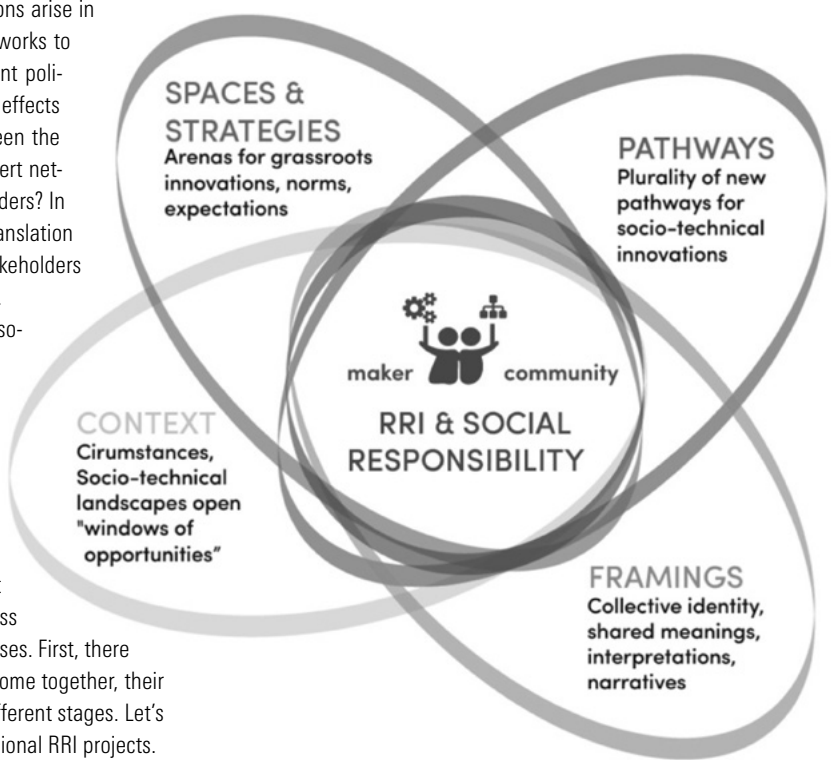
We turn to Michel Callon's Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay (1984) to make clear our use of the concept of translation. In short, the paper describes the scientific and economic controversy about the causes for the decline in the population of scallops in St Brieuc Bay and the attempts by three marine biologists to develop a conservation strategy for that population. Callon treats these attempts as a process of translation. Two insights are relevant for our purposes. First, there is a diversity of actors, and second, as these actors come together, their diversity is being translated, which is a process of different stages. Let's investigate these insights through the lens of the regional RRI projects.

DIVERSITY OF ACTORS

The actors that participate are diverse, as we already noted above. It is important to see these actors as occupying different realms or worlds. What makes for good RRI guidelines in one sphere doesn't make for good RRI practices in another. This is because good RRI is not an external state that is measured, but a social practice of interpretation of the why, how, what, when and with and for whom of responsible research and innovation. Grasping the diversity of stakeholders is a first step in understanding RRI projects as projects of translation. The Critical Making project developed a framework to get a sense of the ways in which stakeholders can be different (see figure 2).

In the model we see four interrelated concepts that may be used to understand who the stakeholders are, where they come from, what they want for the future, and how they want to achieve it. There is the concept of context, which indicates all the historical, political, economic, cultural, religious conditions, as well as other circumstances, issues or situations that matter to the stakeholders and their ability to cocreate. Framings are related to contexts and similarly influential on how stakeholders can participate, yet they determine the stakeholder more specifically. Framings are powerful narratives such as identities or shared meanings. They work as underlying assumptions through social, economic, or political issues, as well as technological frames. Spaces are the arenas that stakeholders want to intervene in meaningfully and these spaces come with different rules and expectations that make specific strategies possible and others not. Spaces can be physical, but obviously social, discursive, or institutional as well. Lastly, the concept of pathways sensitizes us to the plurality of possible ways towards change and innovation. It is useful to keep in mind that there is never one best way, and each way will likely be windy.

Figure 2 Critical Making Baseline model



THE PROCESS OF TRANSLATING THE DIVERSITY OF THE ACTORS

Callon describes the process of translation as unfolding in four stages. The first is the stage of problematization, where researchers enter the scene and define the problem, the solution, and the actors who are part of the arrangement. The second stage is the one of involvement (Callon calls this *interessement*) where researchers try to lock in the actors into the roles of the research protocol. It is not enough for the actors to be identified in the initial stage of the problematization, they must become "interested" in the project, i.e. involve themselves by embodying the roles and relations as defined by the problematization. The third stage is that of enrolment, where things start to move, where the identity of the actors is being tested as they start to negotiate, forge, seduce, consent, or concede into an arrangement, or "multistability", to use the term of Amanatidis and Meijer. Last, there is the stage of representation (Callon calls this mobilization), which designates the process of coming to a characterization of a reality that represents all actors involved. How does this process and language relate to our own regional RRI projects?

START

The first phase combines problem formulation and involvement. Regional RRI projects start with an application process that articulates a problem and hypothesizes around central issues, relevant partners and stakeholders, their interdependencies, and ways to address these issues. The RRI or related regional innovative frameworks that a project subscribes to obviously influences the kinds of partners and stakeholders it seeks out. If citizen engagement is considered crucial for grounding regional RRI, citizen groups should be invited as key stakeholders. If responsible and inclusive engagement is the goal, then co-creation as a method for this makes sense.

Part of this phase is to get stakeholders interested and involved. For this, they need to agree and identify with their role in the problematization, the hypotheses, plan of action and fellow partners and stakeholders. To what extent this involvement really happens is uncertain, as the regional RRI projects reported. Making sure partners and stakeholders fully subscribe to the problematization or the challenge, i.e. take "ownership," is key to successful involvement.

IMPLEMENTATION

The next phase in the process is the execution of the project itself. Within many different work packages, in collaboration with many stakeholders, and in several regions, data collection and analysis are carried out. The focus is on enrolment, on testing the characterization of the issue and the relevance and role of the actors identified. The execution of a complicated research plan that was described in detail in the application often is, as we all know, much messier and unexpected than planned. Things happen (or not) all the time, leading to rearticulations of problems and hypotheses, failing commitments, data that is not useful anymore, drafts of deliverables that are abandoned. Regional RRI needs a lot of work to come into being.

As we described above, engagement of stakeholders is both the central strategy of RRI and its biggest concern at the same time, which needs a lot of attention and time. Selection of stakeholders is often still ongoing and communicating project expectations and RRI benefits to them is not straightforward. Their participation, commitment, use, and time investments become, therefore, unsteady, requiring continuous (re-) negotiation. As we saw earlier, projects reported stories about stakeholders' distrust of project administration, difficulties of being able to relate to the jargon of RRI, resistance to participating in regular assessments, and, lastly, stories about the effects of COVID19 on establishing committed relationships with stakeholders. The systematic involvement of stakeholders continues to be crucial for co-developing the process and taking ownership of the work packages is an important step in this executive stage.

FINISH

The last phase is the representation phase, i.e. the moment where the project is narrativized on behalf of all partners and stakeholders included. Putting together the joint narrative and using suitable indicators is the goal. What this narrative should look like, and which indicators will be used, has often already been promised early on. Moreover, many projects organize indicator development or other evaluative activities throughout the project. It is in the co-implementation phase that these need to be conclusively presented.

What is, however, the 'right' story or indicator? How to account well for the difference between regional innovations? How to account for changes that are more institutional and systemic and take more time than the running time of the projects themselves? Who is the audience? Should narratives and/or indicators represent the collective, the decisions of the project and enable learning, or should they be accountable and convince funding partners or the RRI community of experts? These difficult issues must have been resolved in this last phase, even if thinking about and working towards them has already started in the co-defining phase.

Regional RRI projects are clearly complex endeavors. Callon states that translations are processes of displacements of goals, interests, devices, human-beings, non-human beings, and inscriptions. RRI projects, indeed, aim to displace their allies and make their contexts, framings, strategies and pathways fit within a hypothesis-turned-reality of responsible, regional innovation. The actors don't always behave according to plan of problematization, involvement, enrollment, and representation, however. Research strategies become unsteady, (European) policies and concepts change or disappear, and the funders may raise a brow.

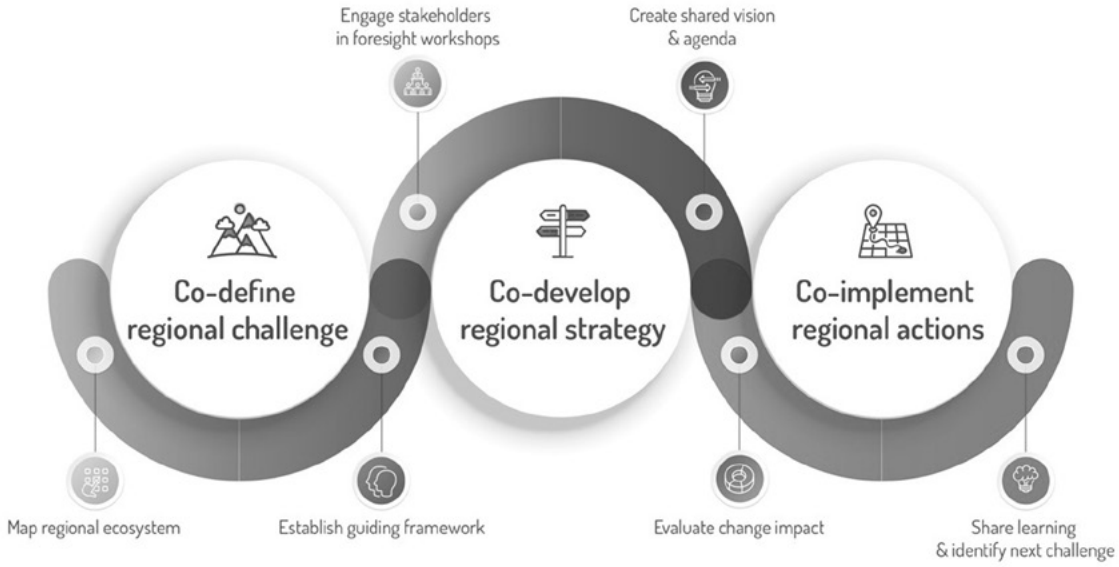
At the end of this process, the project represents the diversity of actors in unison. This process in which actors are defined, associated, and simultaneously obliged to remain faithful to their alliances can be understood as developments in a relationship of power. This development happens in an unsteady way, with many confusions, miscommunications, and misaligned expectations. The resulting translation is therefore not achieved in the most "democratic" way. It is in here that we want to contribute to the quintessentially democratic RRI method of cocreation. While the level of democracy of the process of translation is not Callon's concern, we do think that extra attention to Callon's stage of involvement may offer a way to deal with issues around stakeholder commitment and engagement that regional RRI projects reported.

EVALUATIVE CONVERSATIONS

RRI has been presented in this article as a process in which diverse stakeholders become engaged in a process of regional innovation. This process knows several stages of translation, of becoming engaged together. Maintaining the commitment of these stakeholders is difficult, and this is a crucial condition for the confusions around the different informal and formal uses and needs of assessments and indicators.

There is no general monitoring and evaluation plan to fit all regional RRI purposes as experiences in our regional RRI meetings and the work on multistabilities have shown. The conclusion, therefore, might reach as far as that the "credible contextualization" (one of SUPER-MoRRI's guiding principles) of data and information becomes redundant as each RRI project must have its own, singular M&E framework, and, consequently, project-specific data and indicators. The remaining issue is, then, for each project to establish their own framework for operationalizing RRI and engaging in responsibility and to do this within a collective of very diverse actors.

Figure 3 The SeeRRI model. Created by Nhien Nguyen with graphic design by Marion Magaña



The SeeRRI model. Figure created by Nhien Nguyen with graphic design by Mario Magaña.

We suggest that this requires continuous conversations between all stakeholders. One of our ecosystem partners, the SeeRRI project, designed three stages (SEERRI, 2021) that offer good moments and orientations for th ese conversations. Conversations start in the co-defining phase. This can be followed by negotiations in the co-develop phase, followed by evaluative conversations in later stages. Taking note of RRIs projects struggles with engaging stakeholders and seeing these projects through the lens of Callons translational phases we argue that it is crucial to involve all stakeholders in the conversations and each time again address basic evaluative questions of "where do we come from?", "where do we want to go?", and "what needs to happen?".

Asking these questions allows projects to deal with the mess and uncertainties of research and innovation, thereby overcome the separation between evaluation and RRI. Integrating these evaluative conversations into the project's execution allows formative, real-time evaluation to happen building on co-creation. We are looking forward to experimenting with these evaluative conversations in the future.

REFERENCES

Callon, M. (1984). 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.

European Commission. Horizon 2020 Program. Available: https://ec.europa.eu/programmes/horizon2020/en/h2020-section/responsible-research-innovation

RIPEET (2021). RRI constellation based on RRI dimensions, MoRRI keys and quintuple helix stakeholders. Available: http://ripeet.eu

Schwandt, T. (2015). Evaluation Foundations Revisited. Cultivating a life of the mind for practice. Stanford University Press

SEERRI (2021). The SeeRRI booklet: Bringing responsibility and RRI into regional planning - from theory to practice https://zenodo.org/record/5567217#.YYQ_8S_MzOQ

Sipos, R., Akerman, M. & Wenzelmann, V. (2021). Critical Making Baseline: Towards a Critical Making Responsibility Framework. A report published by the Critical Making Consortium.

Stilgoe, J, Owen, R. & Macnaghten, P. (2013). Developing a framework for responsible innovation, Research Policy, Volume 42, Issue 9, 2013, Pages 1568-1580

SUPER MoRRI (2021). Policy report on the state of SwafS RRI projects. Available: https://super-morri.eu/download/153/findings-and-deliverables/5363/d7-3-policy-report-on-the-state-of-swafs-rri-projects.pdf

Technopolis. (2020). The evolution of Responsible Research and Innovation in Europe. The MoRRI indicators report. Available: https://www.technopolis-group.com/wp-content/uploads/2020/02/The-evolution-of-Responsible-Research-and-Innovation-in-Europe-The-MoRRI-indicators-report-D4.3.pdf

CONTRIBUTORS

SUPER-MoRRI

ANESTIS AMANATIDIS
CWTS Leiden University
Kolffpad 1, 2333 BN Leiden, The Netherlands
E: a.amanatidis@cwts.leidenuniv.nl
ORCID: 0000-0002-8037-7558

TJITSKE HOLTROP
CWTS Leiden University
Kolffpad 1, 2333 BN Leiden, The Netherlands
E: t.j.holtrop@cwts.leidenuniv.nl
ORCID: 0000-0002-1413-8591

INGEBORG MEIJER
CWTS Leiden University
Kolffpad 1, 2333 BN Leiden, The Netherlands
E: i.meijer@cwts.leidenuniv.nl
ORCID: 0000-0003-1481-1739

PAULA OTERO-HERMIDA
INGENIO CSIC-UPV
Camino de Vera, 46022 Valencia, Spain
E: pauother@upvnet.upv.es
ORCID: 0000-0001-6817-3504

SeeRRI

DONATELLA CASALE
University of Haifa
Abba Khoushy Ave 199, Haifa 3498838, Israel
E: dcasale@poli.haifa.ac.il
ORCID: 0000-0001-8758-8588

ALAN HARTMAN
University of Haifa
Abba Khoushy Ave 199, Haifa 3498838, Israel
E: ahartman@is.haifa.ac.il
ORCID: 0000-0002-1275-3902

NHIEN NGUYEN
Nordland Research Institute
Universitetsalleen 11, 8049 Bodø, Norway
E: nng@nforsk.no
ORCID: 0000-0002-6415-467X

TeRRItoria

FABIO FEUDO
K&I
Via Guido Reni, 56 – 00196 Roma, Italy
E: feudo@knowledge-innovation.org

ILDIKO M. IPOLYI
ESF
1, quai Lezay-Marnésia – BP 90015. 67080 Strasbourg Cedex, France,
E: iipolyi@esf.org

TINE RAVN
AU
Bartholins Allé 7, 8000 Aarhus C, Denmark
E: tr@ps.au.dk
ORCID: 0000-0002-8462-5686

TeRRIfica

NORBERT STEINHAUS
Bonn Science Shop
Reuterstr 157, 53113 Bonn, Germany
E: norbert.steinhaus@wilabonn.de
ORCID 0000-0002-6688-0739

SISCODE

ALESSANDRO DESERTI
Politecnico di Milano
Piazza Leonardo da Vinci 32, 20133, Milano MI, Italy
E: alessandro.deserti@polimi.it
ORCID: 0000-0002-1049-7650

FRANCESCA RIZZO
Politecnico di Milano
Piazza Leonardo da Vinci 32, 20133, Milano MI, Italy
E: francesca.rizzo@polimi.it
ORCID: 0000-0001-5809-0684

FELICITAS SCHMITTINGER
Politecnico di Milano
Piazza Leonardo da Vinci 32, 20133, Milano MI, Italy
E: felicitas.schmittinger@polimi.it
ORCID: 0000-0002-8690-2640

CHERRIES

CLAUDIA COLONNELLO
Knowledge & Innovation (K&I)
Via Guido Reni, 56 – 00196 Roma, Italy
E: colonnello@knowledge-innovation.org
ORCID: 0000-0002-2630-2758

DigiTeRRI

MARIANNE HÖRLESBERGER
AIT Austrian Institute of Technology
Giefinggasse 4, 1210 Wien, Austria
E: Marianne.Hoerlesberger@ait.ac.at

RRI2SCALE

CHIARA BUONGIOVANNI
APRE - Agenzia per la Promozione della Ricerca Europea
via Cavour 71, 00184 Roma, Italy
E: buongiovanni@apre.it

RENE WINTJES
Maastricht University
Minderbroedersberg 4-6, 6211 LK Maastricht, Netherlands,
UNU-MERIT
E: r.wintjes@maastrichtuniversity.nl

TRANSFORM

ROGER STRAND
University of Bergen
5007 Bergen, Norway
E: roger.strand@uib.no
ORCID: 0000-0001-6159-1586
THOMAS VÖLKER
University of Bergen
5007 Bergen, Norway
E: thomas.volker@uib.no
ORCID: 0000-0002-2921-3996

TetRRIs

MIKA NIEMINEN
VTT, Technical Research Centre of Finland ltd.
Viikinkaari 1, 00710 Helsinki, Finland
E: mika.niemine@vtt.fi
ORCID: 0000-0001-8528-6869

EMAD YAGHMAEI
Delft University of Technology (TU Delft)
Mekelweg 5, 2628CD Delft, Netherlands
E: e.yaghmaei@tudelft.nl
ORCID: 0000-0003-4884-77801

RIPEET

GABRIELE QUINTI
Knowledge & Innovation (K&I)
Via Guido Reni, 56 – 00196 Roma, Italy
E: quinti@knowledge-innovation.org
ORCID: 0000-0002-8630-5686

KEYWORDS: evaluation practices, RRI, SwafS, translation, stakeholders

METAMORPHOSES AND PERFORMATIVITY. TRANSFORMATIVE R&I POLICIES AND THE NORM(ALIS)ING EFFECT OF SOCIETAL IMPACT

LISE MOAWAD AND CORNELIA SCHENDZIELORZ

DOI: 10.22163/fteval.2022.545

ABSTRACT

In 2014, UK higher education institutions implemented a new system for assessing the quality of research, the Research Excellence Framework (REF) and took the opportunity to introduce "impact beyond academia" as a 'new' assessment criterion. Transformation and innovation-oriented R&I policy are roughly similar in Norway and the Netherlands regarding underlying ideas as well as timing. In occasion of this convergence this article tackles the discursive and performative construction of "societal impact" as a metamorphic constantly changing, transforming, and evolving criterion. Using data from policy documents from the UK, the Netherlands, and Norway from 2014 until now, the comparative semantic analysis draws on theories of speech acts and performativity to reveal the dual effect (normalising and norming) of the discursive device by R&I policymakers. The resulting typology, based on four criteria (terminology, positive and negative valences, *oikonomia* of knowledge and policy slogan), sets the ground for the exploration of further dimensions of societal impact evaluation challenges.

> "In nova fert animus mutatas dicere formas / Corpora."
> Ovid, Metamorphoses (I, 1-2)

INTRODUCTION

In 2014, UK higher education institutions implemented a new system for assessing the quality of research, the Research Excellence Framework (REF), thus replacing the previous Research Assessment Exercise (RAE) and establishing the "*impact beyond academia*" as a 'new' assessment criterion. The very same year, the Dutch government issued a report on research policy, calling for "*maximum impact*" for Dutch science (2025 – Vision for Science, Choices for the future). Still in 2014, the Norwegian Ministry of Education and Research edited a long-term plan for research and higher education 2015–2024, aiming among others to "*tackle major social challenges*" (NMER 2014). A few years later however, discourses took on a different shade of meaning. In the REF2021, impact case studies were worth 25% of the overall profile (an increase from 20% in 2014), where public engagement (reach of impact) and impact on teaching (significance of impact) are also given consideration in the UK assessment system. For its part, the Dutch government encouraged the dialogue between science and society "*by targeted communication and outreach activities*" (NWO - Dutch Research Council). Finally, the Norwegian government claimed in 2018 that "*knowledge development is driven by more than goals and targets*" (NMER 2018).

In all three quoted examples, the formulation and quantification of new objectives confirm that innovation and valorisation policies have an increasing influence on academic practice worldwide (Dance 2013). Moreover, financial R&I instruments are connected with higher demand for regulatory policy instruments (Dinges et al. 2020). Among them are discursive strategies, and above all concepts, which make it possible both to set the rules of the game and to adapt, if the situation requires it. They contain an internal tension (but not an opposition) and are in this sense as fascinating as they are difficult to define. 'Societal impact' is one perfect example of this dynamic tension.

In this article, we tackle this very changing nature of the impact definition from a linguistic perspective by examining the performative dimension of societal impact as a scientific concept: To what extent are the semantic instabilities around this notion an obvious sign of the changing priorities of political stakeholders, between pressing societal challenges and economic development? Using data from multiple data sources from the UK, the Netherlands, and Norway from 2014 until now, the comparative Critical Discourse Analysis (CDA) draws on theories of speech acts and performativity to reveal the dual effect (normalising and norming) of the discursive strategy of 'impact' by R&I policymakers. The resulting typology is based on four criteria: terminology, positive and negative valences, *oikonomia*[1] of knowledge and policy slogan. It sets the ground for the exploration of further dimensions of societal impact evaluation chal-

1 Following the works of Michel Foucault (Foucault 1975) and Giorgio Agamben (Agamben 2011), we decided to prefer the Greek word οἰκονομία (oikonomia) over the contemporary word 'economy'. From the original meaning of the word ("management of a household or family, husbandry, thrift", quoted from Liddell & Scott 1940), we thus emphasise its practice-oriented dimension and can define it as 'a form of arrangement and disposition of the knowledge system (actors, technologies, forms)'.

lenges related to the specific transformation and innovation-oriented R&I policies.

BACKGROUND: EVALUATION CRITERION AND INNOVATION POLICIES

Evaluation in science, and especially the analytical distinction between value - the basic categorisation of persons, objects and practices as valuable or worthless - and values - the normative value systems through which actions can be evaluated as right or wrong -, is already object of a subfield of its own in sociology (Lamont 2012; Krüger & Reinhart 2016). Whether one speaks of the rise of the evaluative state (Neave 1988) or the audit society (Power 1999), the observation is the same for many scholars: audit procedures are redefining accountability, transparency, and good governance in all aspects of society, including the higher education field (Shore & Wright 2015), and research assessment shapes the environment it seeks to control - namely institutional behaviours and organisational cultures (Crawford 2020). The (e)valuation criterion of societal impact, which has been used in the REF since 2014, is a striking example of this regulatory culture. This question is intrinsically tied to the way transformative R&I policies are framed and frame themselves. Considering concepts such as 'innovation for growth', 'national innovation systems' or 'transformative innovation policy' – the latter drawing particular attention to the direction of innovation, i.e., to the social and political choices embedded in technology (see Weber & Rochracher 2012; Schot & Steinmueller 2018; Diercks et al. 2019) - imply that looking for the societal impact might mean establishing an indicator for the innovative power of scientific research for society in procedural terms.

The aim of this paper is not to present a comprehensive review of the existing research on, and practices employed in the assessment of societal impact (see e.g., Bornmann 2013), but rather to point out some typical cases of power dispositives for framing it and thus producing, in deed and word, a "new" world of things (Berger & Luckmann 1966). Whether we talk about incentive policies that encourage societal impact via binding tools such as amendments or legislation (de Jong et al. 2015) or look at procedures, processes and roles through the prism of the Foucauldian apparatus (Wróblewska 2018), this control manifests itself in several ways and levels, both practical (e.g., what evaluation systems do to research, see Hessels & Smit 2021) and rhetorical (Hesselmann & Schendzielorz 2021). The following contribution represents a straight continuation of this shift towards more attention devoted to language and language practices in STS, while investigating the articulation of the different R&I policy discourses around this pattern of 'societal impact' (Foucault 1966) and their political (and therefore productive and serviceable) effectiveness. We then hereafter propose a linguistic analysis as close as possible to the text - an approach rarely taken to this extreme in this field.

THEORY: SPEECH ACTS AND PERFORMATIVE LANGUAGE

The theory of speech acts - acts done in the process of speaking - and their political effects (staging, ordering reality, producing a vision of the world) is based on the idea that language functions as a form of social action that not only has propositional content but is an action on its own through "performative utterance" (Austin 1962; Searle 1969). Moreover, as the speakers' linguistic effectiveness depends on their social authority (Bourdieu 1994), words can be considered to be political in themselves. As ideological vectors and epistemic labels (Foucault 1966), they participate in the production of credible authority by performing legitimacy (Butler 1997); and the instruments in action, far from being neutral auxiliaries, contribute to discursive formatting. The policy documents we will analyse are therefore, in form and substance, nothing more than the realisations (passive and active) of this power. Since words are not only a call for action, but also elicit emotions, this dimension also has to be included in the establishment of analysis criteria (see the Data and Methods section), in particular that of valence (Frijda & Mesquita 1998).

The question of the link between performativity (of political discourse) and productivity/efficiency (of knowledge-producing bodies) is thus at the centre of our reflection. If we believe that the way '*impact*' is defined determines how it is assessed (Donavan 2011), the analysis of political practices around the definition and imposition of societal impact as an evaluation criterion may provide some elements of an answer.

The working hypothesis and its corollaries that we put forward and intend to discuss in this article therefore fall within this dual theoretical framework:

H: Through the definition of 'societal impact', R&I policymakers perform an indicator for the innovative power of science research for society.

C1: This definition in procedural terms normalises *the assessment devices and processes.*

C2: This definition in procedural terms norms *the assessment devices and processes.*

Drawing on Foucault (Foucault 1978), we hereby make the distinction between discursive and non-discursive practices of *normalization* ("What is normal?"), in which assessment devices and procedures are aligned with the currently perceived common ground of research assessment, and practices of *normation* ("What is the norm?"), in which standards are established that function as norms in the sense that they set benchmarks against which assessment devices are measured in the future.

CASE SELECTION AND SAMPLING

The starting point for the analysis was the RAE and the REF, from which the criterion of impact was established as an evaluation criterion in the European research area. As the development, structure and proceedings of the REF has been widely and abundantly studied (a. o. Watermeyer 2014, Watermeyer & Chubb 2019, Wróblewska 2018, Smith et al. 2020), the question arises to what extent the dynamics of the concept "societal impact", formulated in connection with the two above-mentioned British evaluation systems, result in repercussions and effects in other European research systems. We therefore favour

to continue and expand (Wróblewska 2019) the comparative analysis to examine those conceptual and institutional variations and cope with national path-dependencies. Indeed, discourses of funding and innovation agencies vary first and foremost across nations because of language practices. In order to ensure the congruence of the linguistic discourse analysis, we hence selected countries in which English is dominant as lingua franca in research as well as in research policy, assessment and funding. Other endogenous factors deserve to be taken into account, beginning with national cultural scripts (Wierzbicka 1994), i.e., common beliefs (Shepsle 2010), common expectations (Hall & Soskice 2001), or particular elements of national and ideological repertoires (Lieberman 2002). To overcome these challenges, we choose a small-N comparison (Mahoney 2003; Skocpol and Somers 1980) with only three cases under observation: the UK, the Netherlands and Norway, as all three are among the top countries on the European Innovation Scoreboard and with a strong English-speaking research culture. On the one hand, the resulting qualitative analysis will more clearly reveal the historical and political contingencies of the macro-social units studied; on the other hand, the extension of the concept of 'societal impact' will be limited, and the cross-linguistic issues will be more easily traceable.

First of all, the UK is picked as an influential case (Seawright & Gerring 2008), as it has an established and influential performance assessment system that permeates all other European national practices: the country has a long history of performance enhancing instruments (RAE in 1986, REF from 2014 onwards), and the RAE/REF experience has been considered by other EU countries as a success "*in delivering the long-standing science policy goals of government*" (de Boer et al. 2015). Secondly, the Dutch science policy governance system is typical of a desire to strengthen its overall competitiveness in terms of World Economic Forum scores (NIFU 2016). "*It places much emphasis on the commercialisation of public research*" (OECD 2014) and emphasises inclusive deliberation and collaboration (Molen et al. 2019). Finally, Norway is a typical case of a still very national innovation policy based on trust and a close relationship between the government and the higher education institutions. The country has a strong tradition of investing in resource-based sectors, regarded as relevant for societal challenges (OECD 2017).

DATA AND METHODS

Empirically, several types of material will be collected and put into perspective to test the hypotheses formulated above. Based on policy documents (government action plans, institutional websites of funding agencies, joint statements by relevant intermediaries)[2], we develop a systematic synopsis of the range of meanings of this criterion – including linguistic and etymological inscriptions, dispersion of occurrences in the discourse (or collocations, see Halliday 1966) as well as associated horizons of interpretation. The degree of precision of the analysis of the texts varies according to the nature of the documents analysed. Documents such as the public statements of various actors, whether governments, funding agencies or individual politicians were worked through in detail to try to understand the nuances of language precisely. Others, such as evaluation reports commissioned by governments from external agencies, were read and treated more broadly to see which themes are highlighted in the public communication of science policies. Finally, a last group of documents (mainly institutional websites) has undergone several successive analyses in order to understand how central the determination of 'impact' is for self-presentation and what specific weltanschauung is being conveyed. This descriptive investigation sets the ground for a following in-depth analysis through four one-to-one semi-structured expert interviews serving to explore further dimensions of societal impact evaluation challenges related to the specific transformation-oriented R&I policies in UK, Norway, and Netherlands, and helping to fill the interpretive gaps of a political language whose rough edges have been smoothed out. The interviewees are there to help pose the problem, they consolidate the hypotheses, but the following analysis is not primarily based on their answers.

For this purpose, a methodology that explores the relationships between (non-linguistic) social practices and linguistic practices (such as CDA) may be the most insightful (Fairclough 1989). Indeed, unlike other methods, it places particular emphasis on social pressing issues and, in so doing, makes language much less abstract by giving words meanings dependent on the social, economic, and political context in which they are uttered (MacGregor 2010). As this dimension has fairly rarely been the focus of the previous analyses of this concept and its uses, we assumed that such a method was more likely to carry out comparisons of processes, procedures and measures and concrete policy implementation. Using CDA tools, we were then able to combine a qualitative structuring content analysis with a discourse analysis of selected passages (Stamann et al. 2016), where particular attention was attached to the connotative meanings of the notion of 'impact', as it covers all the indirect, peripheral, subjective, cultural, implicit, and other contextual meanings that can be generated by elements of discourse (Trask 2007). After a close-reading analysis of the documents, we coded them according to a set of predefined and ex ante validated criteria, mainly: **terminology** (how concepts are labelled and designated), **positive and negative valences** (the affective quality of the situation, namely the intrinsic "good"-ness or "bad"-ness of the words), ***oikonomia* of knowledge** (in the etymological sense of the term, as a household management practice), **policy slogan** (regimes of repetition, participation and engagement). These are supplemented with case-specific special features as needed. Ultimately, the observations are summarised in a concluding synopsis.

CASE STUDY 1: UNITED KINGDOM

ETYMOLOGY AND TERMINOLOGY (UK)

Returning to the etymology of *impact*, the Oxford English Dictionary traces the word back to the Latin *impactum*, the perfect passive participle of the verb *impingere*, which means "*to dash against, throw on, thrust at, fasten upon*" (Lewis & Short 1879). It refers both literally to "*the act of impinging*", namely "*the striking of one body against another; collision*" (chiefly in Dynamics, in reference to momentum), and figuratively to "the effective action of one thing or person upon another; the effect of such

2 See the list of quoted policy documents and reports in the appendix.

action; influence; impression", particularly in the time-honoured phrase "to make an impact (on)". The sense of "strike forcefully against something" is first recorded 1916, and the figurative sense of "have a forceful effect on" can only be traced from 1935 (Online Etymology Dictionary).

In the field of science and innovation policy, the political actors who instigate the idea of impact assessment and negotiate the concrete meaning of the notion as well as the way it would be assessed are various and numerous in the British context (Wróblewska 2018). Consultations between them resulted in a comprehensive definition of 'impact' understood as (but without being limited to)

> *"an effect on, change or benefit to: the activity, attitude, awareness, behaviour, capacity, opportunity, performance, policy, practice, process or understanding; of an audience, beneficiary, community, constituency, organization or individuals; in any geographic location whether locally, regionally, nationally, or internationally. Impact includes the reduction or prevention of harm, risk, cost or other negative effects" (REF 2011).*

This definition remains in place for REF2021 (REF 2020).

THE USE OF HISTORY AND LAW IN THE INJUNCTION TO CHANGE (UK)

The rhetorical and conceptual framework in which this new terminology is embedded and developed is the 'reform' one. The injunction to change is formulated via rhetorical strategies whose procedural significance can be of great interest. For instance, in the RAND report, the conjunction *because* is used in the same anaphoric way as the visa formulas in the preambles of legal texts (conjunction *whereas*, gerund clauses): "*Because of the diverse nature of impacts*", "*Because of the imperfections of both quantitative and qualitative measures*"... It thus seems to open the statement of a text which, on the model of a legal document, serves as a basis for the decision-making power and the decision-making act (RAND 2010).

The political vocabulary also borrows from the humanities, especially history, to convince and persuade. The rhetorical motif of 'reform' is mainly a facility of language, where change is described in terms of rupture/continuity, a dichotomy that is certainly traditional in political science (e.g., Collier/Collier 1991; Birkland 1998) but still effective when it comes to discursive strategies. The double narrative "*Building on Success*"/"*Learning from Experience*" (Stern Report 2015) is supported by the review of the REF Report (Technopolis 2018) commissioned in 2016 by the British Department of Business, Energy and Industrial Strategy (BEIS), which emphasises on the one hand the REF's place in a global history ("*probably the oldest [performance-based research funding system]*"), and on the other hand its role as an icebreaker and its knock-on effect ("*The REF2014 is arguably the first major discontinuity in the development of the REF [RAE]*"); "*this was seen as a completely new idea*"). The word impact thus retains a double argumentative force: it both arouses support, via an inscription in the past, and desire, via a rhetoric of progress.

POSITIVE AND NEGATIVE VALENCES (UK)

In the UK policy documents analysed for this paper, impact is presented positively in all its meanings (Bornmann 2013): societal products or outputs (new products and services with added value), societal use (societal references), societal benefits (changes in society). The official political discourse is that of a '*success story*' (Stern Report 2015). Even if its percentage is lower than that of excellence in the REF calculation system, it becomes the default evaluation criterion, to the extent that 'non-impact' (outputs that are not exploitable, references that are not productive) is counterintuitively justified in the same terms as impact. However, the Stern Report does not so much deploy a coercive discourse as an inclusive narrative about the participation of policymakers in the creation of favourable conditions for the production and dissemination of knowledge. This euphemisation of the discourse, via verbs such as *contribute*, is thus merely a rhetorical strategy that changes the valence of the description (i.e., the intrinsic affective quality of the situation) from bad (*forces) to good (*helps).

As we can see, *impact* contributes to the establishment of a consensus, the status of which as a social referent can lead to resistance or delegitimisation ("*I'm now some kind of civil servant charged with delivering the government's priorities*" wrote one academic quoted by The *Guardian*, 13.10.2009) or full and complete adherence, or even a bidding war, thus becoming a scholarly distinction (Watermeyer & Chubb 2019). For policymakers, this normalisation is accompanied by the diffusion of an idealised and polished vision of research as the hegemonic norm: goal-oriented, linear, devoid of obstacles, depersonalised and always excellent.

OIKONOMIA OF KNOWLEDGE (UK)

The RAE was introduced in 1986 at a time when Margaret Thatcher wanted to "*get better value for money through greater efficiency*" (Leader's speech, Brighton 1984). A few years later, the discourse accompanying the creation of the REF remains largely influenced by the conventional economy: one of the main motivations to build a new framework for assessing research quality in the UK has been to "*produce robust UK-wide indicators of research excellence for all disciplines which can be used to benchmark quality against international standards and to drive the Council's funding for research*" (HEFCE 2007). Another HEFCE report makes this threat to cut funds explicit: "*The economic landscape in 2009 was very different to what we had experienced over the previous 10 years...The period of growth in public funding enjoyed by HE over the past decade is over and unlikely to return for some time*" (HEFCE, 2010).

Incidentally, the Economic and Social Research Council's definition also encompasses economic performance and competitiveness: "*economic and societal impact, which is the demonstrable contribution that excellent social and economic research has on society and the economy, and its benefits to individuals, organisations or nations*" (UKRI 2021). Finally, as a last illustration of this semantic obsession, we may note the emphasis on directly quantifiable financial impacts from research via the use of a synecdoche which the Russell Group universities boasts in a 2012 paper: "*our definition of 'economic impact' includes social impacts*". Economics (and not just any economics, but the one that marginalises heterodox discourses, see Stockhammer et al. 2021) thus 'represents' the social, understood as political representation: it embodies it, it acts on its behalf; and in both cases, if it makes the other (i.e., the social) present, it is on the condition that it replaces it.

POLICY SLOGAN, MAGICAL FORMULA (UK)

The academic literature on the REF tells the myth of an evaluation criterion created almost *ex nihilo* (Kogan & Hanney 2000; Bandola-Gill & Smith 2021). Such self-narratives are often reconstructed in retrospect. However, the increasing number of occurrences of the term '*impact*' in successive HEFCE annual reports and accounts clearly reveals the inflation of the formula: from seven occurrences in the 2004-2005 report, for example, to 19 (report 2008-2009), then to 36 (report 2010-2011) and 34 (report 2014-2015), reaching its peak in 2016 (43 occurrences in the 2015-2016 report) before deflating back to 27 (report 2017-2018). Even if the use of the term goes beyond the sole notion of "*societal impact*", there is a diffusion by capillarity of the uses of this term which ends up being applied to other contexts. It is thus obvious that while the term was already well established in political discourse in 2014, its use has exploded in the first phase of its life, i.e., the momentum of its problematisation (Wróblewska 2021).

CASE STUDY 2: NETHERLANDS

TERMINOLOGY (NL)

In the Netherlands, there is a real specificity in the national definition and understanding of what societal impact is. In most of the policy documents analysed, the term '*valorisation*' (in Dutch '*valorisatie*') is preferred to '*societal impact*', which is borrowed from English and has more violent connotations, as one of the interviewees explains: "*there was a shitty disaster movie from the 2000s where asteroids hit the earth, and I think it was called Impact or something like that. So I always think of that so I don't use the term*" (Interviewee 2). Overall, the term 'impact' appears to be a linguistic import that has spread beyond British borders without being a pure linguistic translation ("*I don't think it was that easy that the Dutch simply adopted the English term. I think there's more …*", Interviewee 2).

There are many different definitions of '*valorisation*' in circulation, nearly one for each policy agent (Ministry of Education, Ministry of Economy, NWO, VSNU, KNAW). The different aspects are reflected in the definition proposed by the Dutch government in 2009 (quoted by Drooge & Jong 2015):

> *"the process of creating value from knowledge by making knowledge suitable and/or available for economic and/or societal use and translating that knowledge into competitive products, services, processes and entrepreneurial activity."*

Since 2014, the question may have arisen of introducing a new term, although some experts advise against this (Jong 2015): such a definitional shift would take time, both in the upstream design and in the negotiations it would entail.

POSITIVE AND NEGATIVE VALENCES (NL)

Official policy documents make good use of those many terms, especially in their English-language communications. The Dutch Research Agenda also encourages the dissemination of knowledge for a "*positive and structural contribution to the global society of tomorrow*", the idea being "*to build bridges today in order to jointly address the scientific and societal challenges of tomorrow*" (Dutch Research Agenda 2019-2022). The government encourages the co-construction and circulation of knowledge, and open science that is beneficial to society as well as future-oriented. However, it calls for everyone to be vigilant: technological innovation must be accompanied, because "*it cannot be assumed that this impact will necessarily be positive, for which reason it is essential that science and society maintain an ongoing dialogue*" (2025 - Vision for Science, Choices for the future). It is noteworthy that, contrary to the predominantly positive connotation, the concept of '*impact*' is not associated solely with a positive phenomenon.

OIKONOMIA OF KNOWLEDGE (NL)

From the outset, the Dutch government has emphasised the monetisability of public research - making the commercialisation of research the main issue, as evidenced by an OECD report commissioned by the Dutch government (OECD 2014). The report 2025 - Vision for Science, Choices for the future, published in 2014, defines valorisation as the "*use of knowledge to gain some economic advantage, but also its use with a view to solving societal issues or contributing to the public debate*". Vocabularies are clearly economy- and business-oriented: "*Given our culture of cooperation, the Netherlands is extremely adept at finding new combinations and opportunities for cross-pollination*" (Vision for Science 2014). The prepositional phrase (beginning with '*given*') acknowledges the existence of such a culture by definitively assigning a characteristic (the cooperative tendency) to the entire Dutch population - an assignment of identity reinforced by the use of the metonymy '*the Netherlands*'. The business buzzword '*cross-pollination*' serves as a discursive marker to ideologically frame the political thought pattern at work here: knowledge is first and foremost an economic good, which must be treated as such. The impact argument thus becomes a bargaining chip, as explained by one of the interviewees:

> *"I never understood what it means exactly, where it comes from. But what I think it meant for policymakers was how does your academic knowledge help companies make money. It was often used as a kind of a code word for the commercial potential of research."* (Interviewee 2).

POLICY SLOGAN, SUPERLATIVE BUZZWORDS (NL)

As we can see, the Dutch science policy seems to aim at developing closer relations between science, society, and the private sector, all "*with maximum impact*" (2025 - Vision for Science, Choices for the future). The stakes appear to be high for the Dutch government, when one observes the co-occurrences of the concept '*impact*' in this very report on research policy: "*increase the impact of science*", "*maximum impact*", "*the greatest possible impact*", "*huge potential impact*" etc. The accumulation of strong adjectives, even superlatives, indicates the importance that is given, at least on paper, to this dimension. The use of vocabulary with religious connotations ("*particular attention should be devoted to the circulation of knowledge and skills*") also reinforces the impression of a mission assigned to political stakeholders. Impact or valorisation is distinguished by

its incantatory dimension, and the broadening of the extension of these concepts contributes to maintaining the vagueness around them, to the point of making them excessively ductile, or even empty, as one of the interviewees points out:

> *"It's an interesting question how these buzzwords develop. So there are complex discursive processes at play through which certain terms become popular. And often, it's about precisely the fact that they are quite malleable that you can sort of interpret them in different ways. It makes them, you know, useful and practical"* (Interviewee 2).

CASE STUDY 3: NORWAY

TERMINOLOGY (NO)

For many Norwegian researchers, their first encounter with the term was via European science and innovation policy: "*the idea that you should document the potential impact of your research came to them first from the EU system*" (Interviewee 4). In Norway, the English term '*societal impact*' is used, but it is not the only one. The lexical variations mobilised by Norwegian policy stakeholders depend in particular on the disciplinary field, as reported by Wróblewska (2019), who cites the terms '*samfunnsbidrag*' (societal contribution) for the humanities, '*samfunnseffekter*' (societal effects) for the applied sciences or '*samfunnsbetydning*' (societal significance). There is a literal translation of the term (Norwegian: '*virkninger*' or '*effekter*'), but it seems to be little used in 2014. Instead, a foreign terminology from the British REF is preferred, which has two advantages, according to some interviewees:

> *"And then we looked to the British REF for inspiration and we decided to use [...] the REF definition of societal impact because we thought it was quite open to all types of impact. So it would be possible to use it for different disciplines. And also we found it an advantage that it was already known to the research community. So it would be known to the peers that we invite. We always use international peers"* (Interviewee 4).

Overall, the Norwegian definition of impact seems to be much more permeable to supranational discussions on defining major social issues than elsewhere, as evidenced by the mention of the Paris Agreement and the UN 2030 Agenda in the long-term plan for research and higher education 2019-2028.

POSITIVE AND NEGATIVE VALENCES (NO)

Yet, the societal component of research appears from the outset in all the Norwegian policy documents we have been able to consult, albeit in different terms. Solving the "major challenges to society" is one of the three main objectives of the government's long-term plan for research and higher education published in 2014, alongside strengthening competitiveness and innovation capacity and developing high-quality research groups. However, the official political discourse goes beyond the goal-oriented dimension of research, returning to the narrative motif of the researcher-discoverer: "in many cases, it is curiosity-driven research that has led to the most extraordinary results" (Long-term plan 2019-2028). The universe drawn discursively in these policy documents is desirable: in the sentence "[this research] generate[s] knowledge that can give people better, richer lives" (Long term plan 2019-2028), the use of plurivalent qualifying adjectives (good, rich) in degree 1 (comparative) makes it possible to provoke incorporation (Maingueneau 1999) in readers or listeners, i.e., to make them adhere to the universe of meaning proposed to them. The vocabulary is sometimes so meliorative in policy documents that some Norwegian researchers (Sivertsen/Meijer 2019) point to the gap between the government's expectations of research ('extraordinary impact') and the actual results that researchers think they can prove and communicate ('normal impact').

OIKONOMIA OF KNOWLEDGE (NO)

The Norwegian government makes immediate use of the lexical field of the market economy: "*value creation*", "*quality of the workforce and the services delivered*", "*new solutions and products*", "*adaptability and increased productivity*" (Long-term plan 2015-2024). This ideological marking is confirmed by one of the interviewees:

> *"About the specific idea that investments in research should provide societal returns: I think it's very much a part of the whole period of globalization and economic growth that we have"* (Interviewee 4).

If Norway brands itself as a "*knowledge nation*" according to the government's official website, it is because this knowledge and expertise are above all considered to be among their "*most important competitive factors*" (Long-term plan 2015-2024). The ambition is clear: the government announces its goal "*to make Norway one of the most innovative countries in Europe. Like other high-cost countries, Norway's competitive approach must incorporate knowledge as a basis for innovation and higher productivity*" (Long-term plan 2019-2028). Here again, the excessive use of degree 1 (comparative) and 2 (superlative) adjectives is typical of a political discourse that aims to convince as much as to persuade.

Presented as a public good, knowledge is very similar to traditional goods and services. Behind the apparent obviousness of fixed concepts and broad categorisations, it is a peculiar conception of the world that is imposed, via verbs expressing a normative modality ("*Norway's competitive approach must incorporate knowledge as a basis for innovation and higher productivity*") or fixed rhetorical expressions ("*It is therefore important to facilitate renewal and restructuring*" - the author underlines). What is noticeable is that the language of the above-mentioned government action plans is often coercive, in particular when we consider its intentional aim on the receiver, namely its conative function (Jakobson 1960; Austin 1962): this aspect is particularly highlighted in both policy documents, best illustrated by the following performing statement from the Long-term plan 2019-2028: "*It is the Government's ambition to make Norway one of the most innovative countries in Europe*").

A SECTORAL APPROACH TO INNOVATION AND ITS IMPACT (NO)

Most interesting in this case study is the sectoral approach of Norwegian innovation policy, which influences the definition given to societal

impact. The OECD Review of Norway's Innovation Policy highlights this specific institutional configuration combined with a consensus-oriented policy-making style, a principle particularly strong in Norway (OECD 2017). The Norwegian Ministry for Education and Research has by far the largest budget and coordinates policy efforts, along with the Ministry of Trade, Industries and Fisheries and the Ministry for Health and Care Services. Between them, these three ministries account for more than 75% of government allocations for R&D. This sharing of tasks reflects the strategic advantages of Norway around strong industrial clusters and natural resources (climate, energy, medicine, biotech), with a focus on "*global challenges such as climate change, security and preparedness, disease and epidemics, safe access to energy, water and food*" (Long-term plan 2015-2024). Interviewee 4 also reports on this sectorisation of research and innovation policy: "*So we also respond to policy signals from all ministries: when they give money to research, they also have their own priorities*" (Interviewee 4).

The study of the occurrences of the term '*impact*' in the two Long-term plans also illustrates this reduction of the concept's intention. Its economic and sectoral dimension are emphasised: "*significant impact on economic growth, welfare, employment and sustainable development*", "*impact [...] for production of goods and services in the Norwegian private and public sectors*" (Long-term plan 2015-2024), "*impact on the environment and climate*", "*global and local impacts*" (Long-term plan 2019-2028), etc.

POLICY SLOGAN, FUZZY WORD (NO)

Despite the clear neoliberal orientation of the Norwegian government's definition of '*impact*', the evidence of the '*knowledge commissioning - knowledge production - return on investment*' chain underlying the economic definition of the word is questioned by the academic community. Interviewee 4 underlines this: "*It's not a linear relationship, you know, the much criticised linear model that someone is doing research somewhere and then you get some results and then finally something is happening in society*" (Interviewee 4). As an illustration of this difficulty in framing the term '*impact*', the Research Council of Norway refers to the double meaning of this word, which concerns both the "*potential outcomes and impacts of the proposed research and innovation*" and the communication and exploitation part (2021). Putting dissimilar elements - what in stylistics is called zeugme (one concrete, one abstract), on the same functional level, reflects the multiplicity of social and political uses that can be made of '*impact*'.

SYNOPSIS: FLOATING DISCOURSES AND HOW TO HANDLE IT

Many scholars have already established the vagueness of the concept '*impact*' (Watermeyer 2014; Samuel & Derrick 2015; Jong 2015; Wróblewska 2018). But to what extent can we go as far as qualifying '*impact*' as an '*empty signifier*' or at least '*floating discourse*' (Laclau & Mouffe 1985)? These questions about the choice of words are far from being nit-picking discussions, contrary to what interviewees outside of academia might or would like to believe: "*we get a bit too linguistic or semantic, right now...*" (Interviewee 1). Indeed, we have seen how a "*not very theoretically informed*" term (Interviewee 2) can cause confusion and even irritation among receiving parties. What is the point, then, of policymakers having such a diffuse concept?

INTERCHANGEABILITY AND INCREMENTALITY

As we have seen, this instability is partly a step in the process of implementing new research evaluation criteria. Definitions evolve through time; they are works in progress. Indeed, there has been a gradual refinement of definitions in the three countries between 2014 and today.

As a performance management tool, the notion of '*impact*' is both a calculative device that acts as a material '*inscription*' of a managerial construction of reality (Latour 1987) and a '*ritual of verification*' (Power 1999) that represents an '*empty certificate of comfort*' for politicians. In this context, semantic instability is not considered as a problem per se: political stakeholders just have a more everyday and utilitarian use of it: "*If I read documents of the [Dutch] Ministry of Education, for example, they will use valorisation and societal impact, like in the same sentence. They will use it interchangeably*" (Interviewee 3).

TRANSDISCIPLINARITY AND INTERNATIONALISM

The issue of lingua franca in science is a recurrent one. In this particular case, my analysis shows that the fact that a foreign terminology is taken up, as is the case in Norway, can represent two practical interests of policy implementation for the Norwegian political stakeholders: a theoretical one (the term is sufficiently vague for its extension to be encompassing and its field of application to be broad, i.e. transdisciplinary), and a practical one (as the term is already known to the academic community, it will not need to be renegotiated, an element that is all the more necessary since the Norwegian research assessment system makes use of international peers to a certain extent). But if the term 'impact' appears to be a linguistic import that has spread beyond British borders, its connotations may be prohibitive for some academic communities, so that a vernacular word might be preferred, as it is the case in the Dutch system.

RATIONALISATION AND ASSIMILATION

This instability is also part of a political strategy. '*Impact*' may be an empty signifier that can be debated, but it also has the power to integrate and overcome any criticism, as its connotations can be extended endlessly. In this sense, one could even speak of a neutralisation (in the sense of annihilation) of critical or heterodox discourses, as previously mentioned for the UK case, via an economisation of social components of innovation policies.

In practice, the concept of '*impact*' and the related discourses have a unifying force for the community. For example, in Norway, the argument of the benefits of innovation, more than that of impact, is mobilised by the government to justify the principles and methods of rationalisation of the exercise of power. Overall, the inclusive narrative of the co-construction of assessment systems and formats, as developed in the policy

documents of the three countries, is indirectly coercive (see Stern report 2015 for UK), in particular by euphemising injunctions and using rhetorical devices that cannot be countered. In this way, positive valences are evoked by events and situations that cannot fail to win adherence, both collective ("*global challenges*" e.g., in Norwegian policy documents, "*future-oriented*" policies in Dutch ones) or individual (for "*better and richer lives*" in Norwegian political plans).

Noteworthy is to keep in mind that a performative speech of the political stakeholder is above all made possible by the quality of the speaker (his political function, in this case Ministers or high institutional representatives of Higher Education), the recognition of the performativity of the speech by the assembly and the submission to subsequent events, in particular the respect of the commitments made. This makes it all the more understandable that the Norwegian government, for example, is interested in the long-term framework of the Paris Agreement and the UN Agenda.

CONCLUSION AND OUTLOOK

Both normative and exploratory, this analysis should contribute to the understanding of what is at stake, i.e., performed in the definition of '(societal) impact' (that is, beyond academia) from the policy side. The examination of the three cases (the UK, the Netherlands, Norway) demonstrates that societal impact may be considered as a boundary-object (Star & Griesemer 1989). Its interpretative flexibility is the concrete manifestation of changing political priorities concerning research and innovation: by being the subject of definitional bargaining, 'impact' becomes both a rallying and a structuring point for political interests. This article detects and corroborates the double performative effect of bringing to the fore one definition of societal impact rather than another: a normalising effect (through inscription in the past and rhetoric of progress or cross-country linguistic transfers) and a norming effect (through juridification of language, critic assimilation or learning knowledge management techniques).

Nonetheless, it also needs to be asked whether policy makers can simply be characterised as oblivious public servants of economic welfare states economics. The question and pursuit of the usefulness and benefits of science is neither a new nor a specifically neoliberal concern. The struggle for a balance between freedom of research and its limits as well as between truth and utility is an age-old and presumably inconclusive debate (Wilholt 2012; Kaldewey 2013), which finds its respective provisional pacification in more or less equal coexistence of basic and applied research, depending on the epoch and research system. Although assessment systems are also tailored to exploit scientific knowledge production for the national economy, the fact that they stick to innovation discourses at least maintains a narrative that science is essential to society, valuing it positively as a resource and thus enhancing its value. From this perspective, it could even be argued that this rhetoric practices promoting impact are self-motivations and self-persuasions addressing politicians, lay-citizens as well as scientists, which could also open inter- and transdisciplinary research spaces and interrelations enabling not only the exploitation of knowledge for economic purposes but also for the common good. Considering this, the desire for societal impact seems only reasonable as it is in line with the requirement that publicly funded science also justifies itself to society and the general public.

Besides, several things remain to be considered: first of all, we must keep in mind the definitively situated aspect of any performative action. No situation of enunciation can be considered as performative in itself. The research funding system and the weight that traditional political institutions have in it determine the room for manoeuvre of political stakeholders in defining terms. Secondly, there are limitations related to the object of study. Interestingly, this discursive and performative construction of a R&I policy is perhaps even more metamorphic because of the very objects that are studied, namely social transformations and the new rationales and new demands of R&I policymakers related to them. Indeed, such transformations require quick reactivity and high responsiveness from policymakers (Esaiasson & Wlezien 2017) and large resonance from the academic world to reinforce the purpose of political expectations, their accountability, and their accomplishment (Doberneck et al. 2010). Finally, we need to be aware of the bias of comparing native and non-native speakers in their choice of defining vocabulary (Hudson, Detmer & Brown 1995), at the risk of succumbing to conceptual and terminological ethnocentrism. Indeed, comparison of languages relies crucially on the concepts that can be coded with similar effort in all languages - so cross-linguistic regularities should not be forced, overrated, or minimised. Knowing moreover that genuine intuitions about semantic references vary not only across, but also within language cultures, this paper, as a simple conceptual contribution, will have to be complemented by further analyses of this aspect - via, for instance, a broader set of stakeholders' interviews or a broader set of analysed countries.

ACKNOWLEDGMENTS

This analysis is part of the Project "Diversity and adjustability of peer review. On the metastability of peer review formats" (DivA) (2020-2023) funded by the German Federal Ministry of Education and Research (BMBF). For more information, see: https://www.rmz.hu-berlin.de/en/research/Diversity%20and%20Adjustability%20of%20Peer%20Review/diversity-and-adjustability-of-peer-review-on-the-metastability-of-peer-review-formats?set_language=en

We thank all participants of the RMZ Paper Workshop (Humboldt-University of Berlin, Robert K. Merton Center, October 2021) for helping improve and clarify this manuscript, and the two anonymous reviewers for their suggestions and valuable comments.

QUOTED POLICY DOCUMENTS AND REPORTS

HEFCE (2005). Higher Education Funding Council for England account 2004-2005. Bristol: HEFCE.

HEFCE (2007). Higher Education Funding Council for England annual report and accounts 2006 to 2007. Bristol: HEFCE.

HEFCE (2009). Higher Education Funding Council for England annual report and accounts 2008 to 2009. Bristol: HEFCE.

HEFCE (2010). Higher Education Funding Council for England annual report and accounts 2009 to 2010. Bristol: HEFCE.

HEFCE (2011). Higher Education Funding Council for England annual report and accounts 2010 to 2011. Bristol: HEFCE.

HEFCE (2015). HEFCE annual report and accounts 2014 to 2015. Bristol: HEFCE.

HEFCE (2016). Higher Education Funding Council for England annual report and accounts 2015 to 2016. Bristol: HEFCE.

HEFCE (2018). HEFCE annual report and accounts 2017 to 2018. Bristol: HEFCE.

Ministry of Education and Research (Norway) (2014). Long-term plan for research and higher education 2015–2024, Meld. St. 7 (2014–2015) Report to the Storting (white paper). Recommendation of 3 October 2014 from the Ministry of Education and Research, approved in the Council of State on the same day (White paper from the Solberg Government).

Ministry of Education and Research (Norway) (2018). Long-term plan for research and higher education 2019–2028. Meld. St. 4 (2018–2019) Report to the Storting (white paper). Recommendation of 5 October 2018 from the Ministry of Education and Research, approved in the Council of State on the same day (White paper from the Solberg Government).

OCDE (2017). OECD Reviews of Innovation Policy: Norway 2017, OECD Reviews of Innovation Policy. Éditions OCDE. URL: https://doi.org/10.1787/9789264277960-en.

RAND, P. I. (2015). The Policy Institute at King's College London and RAND Europe (2015). Characteristics of high-performing research units: A preliminary analysis Research Report 2015/02 Prepared for the Higher Education Funding Council for England (HEFCE). London: King's College London.

REF (2011). REF2014 Assessment framework and guidance on submissions. Bristol: HEFCE.

REF (2020). Index of revisions to the 'Guidance on submissions' (2019/01). Bristol: HEFCE.

Russell Group of Universities (2012). The social impact of research conducted in Russell Group universities. Russell Group Papers – Issue 3. URL: https://russellgroup.ac.uk/media/5235/socialimpactofresearch.pdf Stern, N. (2016). Building on Success and Leaning from Experience: An Independent Review of the Research Excellence Framework. London: Department for Business, Energy and Industrial Strategy.

Technopolis (2018). Review of the Research Excellence Framework. Evidence Report. London: Department for Business, Energy and Industrial Strategy.

Thatcher, M. (1984). Leader's speech, Brighton 1984. URL: http://www.britishpoliticalspeech.org/speech-archive.htm?speech=130.

UKRI (2021). Defining impact. URL: https://www.ukri.org/councils/esrc/impact-toolkit-for-economic-and-social-sciences/defining-impact/.

United Nations (2015). Paris Agreement (Dec. 13, 2015), in UNFCCC, COP Report No. 21, Addenum, at 21, U.N. Doc. FCCC/CP/2015/10/Add, 1 (Jan. 29, 2016).

United Nations (2015). Resolution adopted by the General Assembly on 25 September 2015, Transforming our world: the 2030 Agenda for Sustainable Development.

Dutch Cabinet (2014). 2025 – Vision for Science, Choices for the future, Wetenschapsvisie 2025. URL: https://www.government.nl/binaries/government/documents/reports/2014/12/08/2025-vision-for-science-choices-for-the-future/visie-wetenschap-eng-web.pdf.

REFERENCES

Agamben, G. (2011). The Kingdom and the Glory: For a Theological Genealogy of Economy and Government. Translated in English by Lorenzo Chiesa (with Matteo Mandarini). Stanford University Press.

Austin, J. L. (1962). How to Do Things with Words. Clarendon Press.
Ball, St. J. (2003). The teacher's soul and the terrors of performativity. Journal of Education Policy, 18(2), 215-228.

Bandola-Gill, J. & Smith, K. E. (2021). Governing by narratives: REF impact case studies and restrictive storytelling in performance measurement, Studies in Higher Education. DOI: 10.1080/03075079.2021.1978965

Berger, P. & Luckmann, Th. (1966). The social construction of reality. A treatise in the sociology of knowledge. Penguin Books.

Birkland, T. A. (1998). Focusing Events, Mobilization, and Agenda Setting. Journal of Public Policy, 18(1), 53-74.

Boer, H. F. (de), Jongbloed, B. W. A., Benneworth, P. S., Cremonini, L., Kolster, R., Kottmann, A., Lemmens-Krug, K., & Vossensteyn, J. J. (2015). Performance-based funding and performance agreements in fourteen higher education systems. Center for Higher Education Policy Studies (CHEPS).

Bornmann, L. (2013). What is societal impact of research and how can it be assessed? a literature survey. Journal of the Association for Information Science and Technology, 64(2), 217-233.

Bourdieu, P. (1994). Raisons pratiques. Seuil.

Butler, J. (1997). Excitable Speech. Routledge.

Collier, R. B & Collier, D. (1991). Shaping the Political Arena: Critical Junctures, the Labor Movement, and the Regime Dynamics in Latin America. Princeton University Press.

Crawford, A. (2020). Societal Impact as 'Rituals of Verification' and The Co-Production of Knowledge, The British Journal of Criminology, 60(3), 493–518.

Crozier, M. (1970). La Société bloquée. Le Seuil.

Dance, A. (2013). Pack a punch. Nature, 502, 397-398.

Samuel, G. N. & Derrick, G. E. (2015). Societal impact evaluation: Exploring evaluator perceptions of the characterization of impact under the REF2014. Research Evaluation, 24(3), 229–241.

Dinges, M., Meyer, S. & Brodnik, Ch. (2020). Key Elements of Evaluation Frameworks for Transformative R&I Programmes in Europe. fteval Journal for Research and Technology Policy Evaluation, 51, 26-40.

Diercks, G., Larsen, H. & Steward, F. (2019). Transformative innovation policy: Addressing variety in an emerging policy paradigm. Research Policy, 48(4), 880-894.

Doberneck, D. M., Glass, Ch. R. & Schweitzer, J. (2010). From Rhetoric to Reality: A Typology of Publically Engaged Scholarship. Journal of Higher Education Outreach and Engagement, 14(4), 5-35.

Donovan, Cl. (2011). State of the art in assessing research impact: introduction to a special issue. Research Evaluation, 20(3), 175–179.

Drooge, L. van & Jong, St. de (2015). Valorisatie: onderzoekers dan al veel meer dan ze denken - e-publicatie met voorbeelden en handvatten om zelf valorisatie te organiseren. Rathenau Instituut.

Ebbinghaus, B. (2006). When less is more: selection problems in large-N and small-N cross-national comparisons. In K.-S. Rehberg (Ed.), Soziale Ungleichheit, kulturelle Unterschiede: Verhandlungen des 32. Kongresses der Deutschen Gesellschaft für Soziologie in München (pp. 4013-4021). Campus Verlag.

Esaiasson, P. & Wlezien, Ch. (2017). Advances in the Study of Democratic Responsiveness: An Introduction. Comparative Political Studies, 50(6), 699–710.

Fairclough, N. (1989). Language and Power. Harlow Longman.

Flick. S. & Herold, A. (2021). Zur Kritik der partizipativen Forschung. Forschungspraxis im Spiegel der Kritischen Theorie. Beltz Juventa.

Foucault, M. (1966). Les Mots et les Choses : Une archéologie des sciences humaines. Gallimard.

Foucault, M. (1975). Surveiller et Punir. Naissance de la prison. Gallimard.

Foucault, M. (1978). Sécurité, Territoire, Population. Cours au Collège de France 1977-1978. Seuil.

Foucault, M. (2004). Naissance de la biopolitique - Cours au collège de France (1978-1979). EHESS/Gallimard/Seuil.

Frijda N.H. & Mesquita B. (1998). The Analysis of Emotions. In M. F. Mascolo M.F. & S. Griffin S. (Eds.), What Develops in Emotional Development?. Emotions, Personality, and Psychotherapy. Springer.

Halliday, M. A. K. (1966). Lexis as a Linguistic Level. Journal of Linguistics, 2(1), 57–67.

Hesselmann, F., & Schendzielorz, C. (2021). Rhetorical power in evaluations: tracing the construction of value-measurement links in debates on societal impact. In P. Dahler-Larsen (Ed.), A Research Agenda for Evaluation (pp. 209-224). Edward Elgar.

Hooghe, L. & Marks, G. (2009). Does Efficiency Shape the Territorial Structure of Government? Annual Review of Political Science, 12(1), 225-241.

Hudson, Th., Brown, J. D. & Detmer, E. (1995). Developing Prototypic Measures of Cross-cultural Pragmatics. University of Hawai'i at Manoa, Second Language Teaching [and] Curriculum Center.

Jacobson, R. (1960), Closing statements: Linguistics and Poetics, Style in language. T.A. Sebeok.

Joly, P.-B., Matt, M. & Robinson, D.K.R. (2019). Research Impact Assessment: From ex-post to real-time assessment. fteval Journal for Research and Technology Policy Evaluation, 47, 35-40.

Jong, St. de (2015). Engaging scientists: organising valorisation in the Netherlands. Doctoral Thesis, Issue Date: 2015-09-10. URL: https://scholarlypublications.universiteitleiden.nl/handle/1887/35123.

Jong, St. de, Smit, J., & van Drooge, L. (2015). Scientists' response to societal impact policies: A policy paradox. Science and Public Policy, 43(1), 102-114.

Kaldewey, D. (2013). Wahrheit und Nützlichkeit. Selbstbeschreibungen der Wissenschaft zwischen Autonomie und gesellschaftlicher Relevanz. Transcript.

Kogan, M. & Hanney, St. (2000). Reforming Higher Education. Jessica Kingsley Publishers.

Krueger, A. K & Reinhart, M. (2016). Wert, Werte, (Be)Wertungen - Eine erste begriffs- und prozesstheoretische Sondierung der aktuellen Soziologie der Bewertung. Berliner Journal für Soziologie.

Kuhlmann, S. & Rip, A. (2014). The Challenge of Addressing Grand Challenges. A Think Piece on How Innovation Can Be Driven Towards the "Grand Challenges" As Defined Under the European Union Framework Programme Horizon 2020. European Research and Innovation Area Board (ERIAB).

Laclau, E., & Mouffe, C. (1985). Hegemony and Socialist Strategy: Towards a Radical Democratic Politics. Verso.

Lamont, M. (2012). Toward a comparative sociology of valuation and evaluation. Annual Review of Sociology, 38(1), 201–221.

Latour, Br. (1987). Science in Action: How to Follow Scientists and Engineers through Society. Harvard University Press.

Lewis, Ch. T. PhD. & Short, Ch. LL.D. (1879). A Latin Dictionary. Founded on Andrews' edition of Freund's Latin dictionary. revised, enlarged, and in great part rewritten by. Charlton T. Lewis, Ph.D. and. Charles Short, LL.D. Clarendon Press.

Liddell, H.G. & Scott, R. (1940). A Greek-English Lexicon. revised and augmented throughout by. Sir Henry Stuart Jones. with the assistance of. Roderick McKenzie. Clarendon Press.

Lieberman, R. C. (2002). Ideas, institutions, and political order: Explaining political change. American Political Science Review, 96(4), 697-712.

McGregor, S.L.T. (2010). Critical Discourse Analysis: A Primer. Halifax. Mount Saint Vincent University

Maingueneau, D. (1999). Ethos, scénographie, incorporation. In R. Amossy (Ed.), Images de soi dans le discours. La construction de l'ethos (pp. 75-101). Delachaux et Niestlé.

Martin, B. R. (2011). The Research Excellence Framework and the 'impact agenda': are we creating a Frankenstein monster? Research Evaluation, 20(3), 247–254.

Molen, F. van der, Ludwig, D., Consoli, L. & Zwart, H. (2019). Global challenges, Dutch solutions? The shape of responsibility in Dutch science and technology policies. Journal of Responsible Innovation, 6(3), 340-345.

Power, M. (1999). The Audit Society: Rituals of Verification. Oxford University Press (Second ed.).

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.

Searle, J. R. (1969). Speech Acts: An Essay in the Philosophy of Language. Cambridge University Press.

Seawright, J. & Gerring, J. (2008). Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options. Political Research Quarterly, 61(2), 294–308.

Shepsle, K. A. (2010). Analysing politics: rationality, behavior, and institutions. W. W. Norton.

Shore, C. & Wright, S. (2015). Audit culture revisited: Rankings, ratings, and the reassembling of society. Current Anthropology, 56, 421–444.

Sivertsen, G. & Meijer, I. (2019). Normal versus extraordinary societal impact: how to understand, evaluate, and improve research activities in their relations to society? Research Evaluation, 29, 66-70.

Smit, J. P. & Hessels, L K. (2021). The production of scientific and societal value in research evaluation: a review of societal impact assessment methods, Research Evaluation, rvab002, https://doi.org/10.1093/reseval/rvab002.

Smith, K., Bandola-Gill, J., Meer, N., Stewart, E., & Watermeyer, R. (2020). The Impact Agenda: Controversies, Consequences and Challenges. Policy Press.

Spaapen, J. & Drooge, L. van (2011). Introducing 'productive interactions' in social impact assessment. Research Evaluation, 20(3), 211–218.

Star, S. & Griesemer, J. (1989). "Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39". Social Studies of Science, 19 (3), 387–420.

Stockhammer, E., Dammerer, Q. & Kapur, S. (2021). The Research Excellence Framework 2014, journal ratings and the marginalisation of heterodox economics. Cambridge Journal of Economics, 45(2), 243–269.

Trask, R. L. (2007). Language and linguistics: the key concepts. Peter Stockwell (Second ed.).

Weber, K. M. & Rohracher, H. (2012). Legitimizing research, technology and innovation policies for transformative change: Combining insights from innovation systems and multi-level perspective in a comprehensive 'failures' framework. Research Policy, 41(6), 1037-1047.

Watermeyer, R. (2014). Issues in the articulation of 'impact': the responses of UK academics to 'impact' as a new measure of research assessment. Studies in Higher Education, 39(2), 359-377.

Watermeyer, R. & Chubb, J. (2019). Evaluating 'impact' in the UK's Research Excellence Framework (REF): liminality, looseness and new modalities of scholarly distinction. Studies in Higher Education, 44(9), 1554-1566.

Wierzbicka, A. (1994). 'Cultural Scripts': A semantic approach to cultural analysis and cross-cultural communication. Pragmatics and Language Learning Monograph Series, 5, 12-35.

Wilholt, T. (2012). Die Freiheit der Forschung: Begründungen und Begrenzungen, Suhrkamp.

Wróblewska, M. N. (2018). The making of the Impact Agenda. A study in discourse and governmentality. Unpublished doctoral dissertation. Warwick University.

Wróblewska, M. N. (2019). Impact evaluation in Norway and in the UK: A comparative study, based on REF 2014 and Humeval 2015-2017. (ENRESSH working paper; Vol. 2019, No. 01).

Wróblewska, M.N. (2021). Research impact evaluation and academic discourse. Humanities and Social Sciences Communications, 8(58), https://doi.org/10.1057/s41599-021-00727-8.

AUTHORS

LISE MOAWAD
Humboldt-Universität zu Berlin
Unter den Linden 6
10117 Berlin
E: lise.moawad@hu-berlin.de
ORCID: 0000-0001-8853-4688

CORNELIA SCHENDZIELORZ
Humboldt-Universität zu Berlin
Unter den Linden 6
10117 Berlin
E: schendzc@hu-berlin.de
ORCID: 0000-0002-4604-5736

KEYWORDS: Assessment systems and formats, societal impact, performativity, conceptual and institutional variations, Critical Discourse Analysis (CDA)

MULTIDIMENSIONALITY THROUGH SELF-EVALUATION:
FROM THEORY TO PRACTICE IN THE BRAZILIAN GRADUATE SYSTEM

ANDRÉ BRASIL

DOI: 10.22163/fteval.2022.546

ABSTRACT

Nearly all science and technology research in Brazil is conducted within a national system of graduate education. Since the 1970s, a graduate program assessment has been an integral part of such a system, and it is currently held on a quadrennial basis. The evaluation model is dynamic, evolving from the experiences of evaluators, policymakers, and the scientific community during each four-year cycle. This study analyses policy initiatives from the 2017-2021 evolving effort, focusing on strategies and recommendations to implement multidimensionality and self-evaluation as integral components of Brazilian evaluation. The paper traces how the idea for a multidimensional assessment was introduced in the country and how U-Multirank, an international ranking of higher education institutions (HEI), has come to inspire an evaluation that is not institutional but of graduate programs instead. The study identified some benefits and limitations of the chosen inspiration and analysed how the Brazilian proposal aligned with the U-Multirank principles. Furthermore, the investigation shows there is little concrete difference from the proposed new model to the one Brazil has already in place. Finally, the last section of this study looks into the once pivotal idea to pursue a self-evaluation component, now relegated to a minor role in the model, but that could be raised to a position supporting the design of an actual multidimensional assessment model.

1 INTRODUCTION

The Brazilian science system is relatively young. By 1965, the country was yet to develop a research tradition, and the little science conducted was essentially confined to a few research institutes and a graduate system of only 27 master's and 11 doctoral courses. Considering that Brazil reached a population of 90 million before the end of that decade, the numbers were far from optimal. However, the scenario started to change over the following years as a robust National System of Graduate Education (SNPG) was launched by a series of government initiatives. This system was conceived based on the core idea that science and education should be strongly connected. As a consequence, most of the country's science and technology research is conducted within graduate programs, both at the master's and doctoral levels (Balbachevsky, 2005; Brasil, 2020; CFE, 1965; Martins, 2018).

Since its conception, evaluation has been an integral part of the SNPG. For nearly 50 years, the Brazilian Agency for Support and Evaluation of Graduate Education (CAPES) – a public foundation linked to the Ministry of Education – has been in charge of evaluating such a system. The adopted model has evolved over the years, and today it includes both a mandatory accreditation and a quadrennial evaluation of graduate programs (PPG). A grading system on a 1 to 7 scale applies, with grades one and two insufficient for accreditation renewal. Superior grades guarantee not only program continuity but may lead to additional funding, access to a broader set of grants, added institutional prestige, and more (Brasil, 2020; Ferreira and Moreira, 2002).

In 2017, the national evaluation assessed the 2013–2016 performance of 4175 graduate programs, with a total of 6303 doctoral and master's courses. The effort involved nearly 2000 panel members, organised in 49 disciplinary committees, in a large-scale endeavour to combine qualitative and quantitative methods to assess the whole SNPG (CAPES, 2018a).

As the evaluation results were announced, Faljoni-Alario et al. (2018) formulated a report with critical suggestions to improve the evaluation model for the 2017–2020 cycle and beyond. The document resulted from discussions between CAPES – as the agency in charge of the evaluation – and coordinators from the disciplinary committees. The report recognised the accomplishments of the evaluation system and included recommendations regarding: time between evaluations, methods for assessing research outputs (e.g., journal articles, books, technical production), criteria and indicators adopted across disciplines, and more. A series of initiatives followed, including the establishment of thematic working groups to propose changes to various aspects of the evaluation system (CAPES, 2019; Monteiro et al., 2019).

Aiming to promote a collaborative redesign of its evaluation model, CAPES organised international seminars to further working group proposals, thus engaging Brazilian academia and the broader scientific community in the debate. As a result, two core concerns became central for the intended evolution of the evaluation system. The first was the need to design a self-evaluation strategy for the SNPG, as its significant expansion had become an obstacle for a central assessment to capture the complex narratives from thousands of graduate programs (Monteiro et al., 2019; Verhine et al., 2019). The second came from the perception that evaluation promoted an overly homogeneous science system since its one-dimensional approach led graduate programs to become substandard photocopies to the top-performing ones. A multidimensional assessment, capable of recognising and valuing differences, was now required (PNPG Committee, 2018; FOPROP, 2018).

This paper investigates ongoing institutional efforts to implement multidimensionality and self-evaluation as components of the Brazilian national evaluation model. Through the analysis of policy documents, legislation, reports, and assessment guidelines, the study traces the motivation and the path towards a multidimensional evaluation, including an overview of the leading proposal for its implementation. The paper also explores the use of U-Multirank – an international ranking of higher education institutions (HEI) – as a source of inspiration for the proposed model, highlighting both the benefits and drawbacks of such adoption. Finally, the study explores the self-evaluation component and identifies the current recommendations for its adoption underestimate its potential to enable a genuinely multidimensional model.

2 TOWARDS A MULTIDIMENSIONAL EVALUATION

The report produced by Faljoni-Alario et al. (2018) was a significant yet initial analysis of what was needed to evolve the evaluation of graduate programs conducted at the time. Considering the need for a broader perspective, CAPES tasked the special committee in charge of monitoring the National Plan for Graduate Education (PNPG)[1] to supplement the material. For that, the group reached out to over a dozen influential organisations in the country's science system, including the Brazilian Society for the Advancement of Science (SBPC), the Brazilian Academy of Sciences (ABC), and the National Council for Scientific and Technological Development (CNPq) (PNPG Committee, 2018, p. 3).

Many of the submitted contributions conveyed concerns about the role of the current evaluation model in shaping a science system that was too academic, focused on the training of future professors for the country's higher education system. A document prepared by the National Forum of Pro-Rectors for Research and Graduate Education (FOPROP) – one of the leading interlocutors between HEI, science policymakers and funding agencies in Brazil – clearly expressed the collective expectations, stating that graduate programs should not be required to excel in every dimension; they could be excellent according to their vocation or specific mission. Evaluation should be able to recognise value across multiple dimensions (FOPROP, 2018, p. 2).

From the joint effort, the PNPG Committee (2018) prepared a report delineating an evaluation model in five dimensions: Training of human resources; Internationalisation; Scientific production; Innovation and knowledge transfer; Economic and social impact. CAPES' Higher Council unanimously approved the proposal, making multidimensionality a priority to evolve the evaluation model (Audy, 2020).

While Audy (2020) mentions the initial multidimensional proposal was not based on any existing system, the aforementioned FOPROP (2018) document suggested the Times Higher Education Ranking as a possible inspiration. Even though the specific suggestion did not seem to find space in the following discussions, it might have directed attention towards other rankings, leading to the discovery of U-Multirank (UMR) as a potential reference to build the new evaluation model.[2]

2.1 U-MULTIRANK: A PROVISIONAL INSPIRATION

The predominant view from many scholars such as van Raan (2005), Calero-Medina et al. (2008), and Gadd et al. (2021) seems to be that rankings are an undeniable part of the higher education landscape with recognised applications despite their evident flaws, biases, and shortcomings. While the objective of this study is not to analyse the value of such rankings, previous research provides relevant arguments to frame their potential as an inspiration to reform a complex national evaluation system such as the Brazilian one.

Starting from the work of Hazelkorn and Gibson (2017), we understand that global rankings often do not count with meaningful, reliable and verifiable indicators and data, especially for international comparisons. As a result, they usually give preferential weight to research outputs, favouring higher education institutions with a focus on the physical, life and medical sciences, and favouring countries where English is the native language. According to Waaijer (2018), that problem is made worse by the fact that most university rankings yield composite scores, often the result of nontransparent raw data, transformation of scores, and weighting. That makes it difficult to analyse the meaningful differences that exist between universities.

In line with the presented perspective, Gadd et al. (2021, p. 16) call for "open and transparent assessment of the relative strengths and weaknesses of the global university rankings to make them more accountable to the higher education communities being assessed". Aligned with that, Moed (2017) mentions current rankings are mostly one-dimensional, and changing that is not a simple task, as even the local, national or international orientation of universities is a dimension often challenging to consider (Calero-Medina et al., 2008).

Considering the presented flaws of global rankings and more, van Vught et al. (2012) introduced U-Multirank as a new approach to ranking in higher education and research. Following a feasibility study conducted by a consortium of universities and research organizations (known as CHERPA), UMR published its first set of results in 2014, aiming to be a "multidimensional", "multilevel", "participative", and "user-driven" ranking.

Fanelli (2016), Hazelkorn and Gibson (2017), and Moed (2017) are among those that recognise UMR to be unlike most international rankings, primarily because of the multidimensional perspective that comes from addressing more than research, as four other essential dimensions of higher education are included: teaching and learning, knowledge transfer, internationalisation, and regional engagement.

UMR's multilevel perspective is about providing information of value for distinct groups of stakeholders. While for some, reports about a particular field may be desirable (e.g., potential students), for others, the

1 See Brasil (2020) for further discussion on the National Plans for Graduate Education.

2 After a preliminary investigation about U-Multirank, a Brazilian delegation visited lead partners of the consortium engaged in its development, in Germany and the Netherlands. Demonstrating the country's commitment to a multidimentional evaluation and U-Multirank influence, the mission included influential representatives of the SNPG, such as the national evaluation director (CAPES), the president emeritus of SBPC, and the president of the PNPG Committee. A month later, UMR representatives also visited CAPES to discuss a potential collaboration (F. Marques, 2019).

institutional-level ranking results might be the most relevant (e.g., HEI managers). Thus, UMR organises its data and indicators aiming to allow for comparisons at the organisation level, but also at the level of disciplinary or multidisciplinary fields (Federkeil, Kaiser, et al., 2012).

For the participative aspect of U-Multirank, the current methodology is based not only on national datasets and collection of organisational or bibliographic data but also on institutional and student surveys conducted for every new edition (U-Multirank, 2021). Finally, for the user-driven perspective, van Vught et al. (2012, p. 3) state that usual university rankings have the pretension of being guided by a nonexistent theory of the quality of higher education, and thus they present collections of indicators as a reflection of a definitive quality of the institution. U-Multirank was designed to be interactive so that users could have control over the available indicators. Information is made transparent so that personal rankings can be tailored to suit specific purposes and users' needs.

Figure 1 shows how this transparency materialises in the ranking results presented in the 2021 edition of UMR. The example retrieved from U-Multirank (2021) shows the "research" dimension and the accompanying indicators[3] for the University of São Paulo (USP), one of the largest HEI in Latin America.

Figure 1: U-Multirank research dimension for the University of São Paulo (USP)

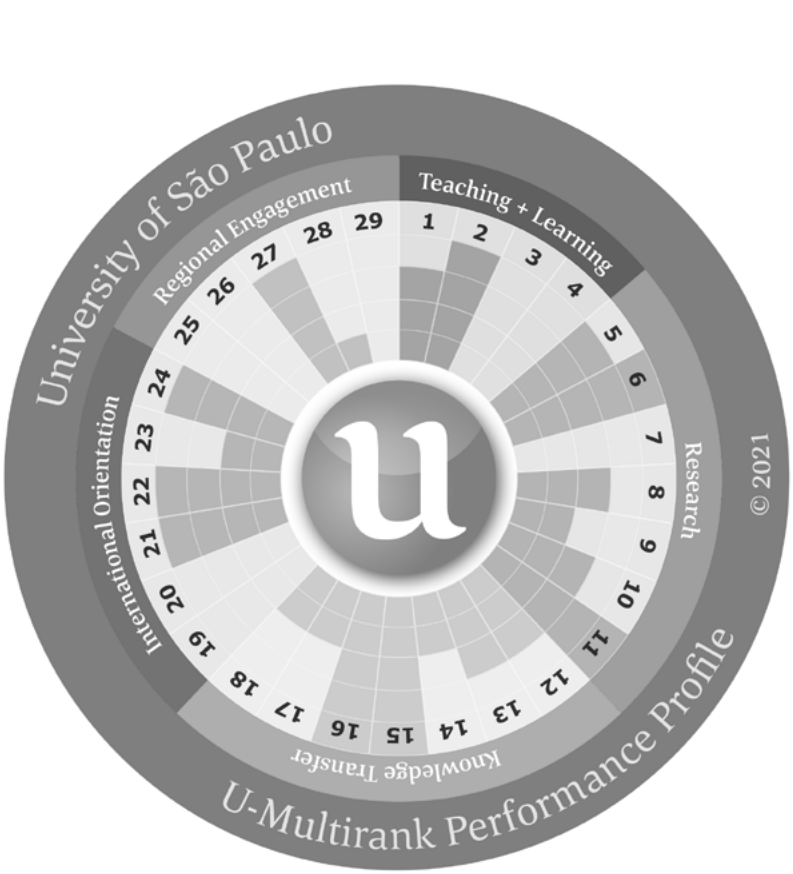
Research			
	Score	*Grade*	*Other universities*
Citation rate	0,77	**C**	0,08 – 2,53
Research publications (absolute numbers)	41.982	**A**	1 – 67.252
Research publications (size-normalised)	0,519	**A**	0 – 13,53
External research income	46,99	**B**	0 – 282,88
Art related output	—		0 – 1,99
Top cited publications	6,5%	**D**	0 – 34,2
Interdisciplinary publications	9,6	**B**	0 – 24,6
Post-doc positions	69,29%	**A**	0 – 100
Professional publications	—		0 – 60,62
Open Access Publications	28%	**A**	

3 In the U-Multirank website, the term "measure" is used to refer to indicators, which will be used in this paper instead, as we consider it to be more appropriate to reflect the type of information included in UMR.

The "research" dimension displayed on Figure 1 includes ten indicators. Calculated scores for the institution are presented against those for all other universities in the database. Data points for USP are shown resized according to the obtained grade in each measure, attributed on a five-level scale: A (very good), B (good), C (average), D (below average), and E (weak). Missing grades are indicated with "–", usually a result of unavailable data from the institution.

The current version of U-Multirank includes 35 indicators distributed across the five dimensions mentioned. A sunburst chart presents 29 of those indicators in a visual profile for each covered institution. An example is shown in Figure 2, once again with data from USP, according to the ranking's 2021 edition.

Figure 2: U-Multirank university performance chart for the University of São Paulo (USP), also listing the 29 indicators displayed in the sunburst



1	Bachelor graduation rate
2	Masters graduation rate
3	Graduating on time (bachelors)
4	Graduating on time (masters)
5	External research income
6	Research publications (size-normalised)
7	Art related output
8	Citation rate
9	Top cited publications
10	Interdisciplinary publications
11	Post-doc positions
12	Income from private sources
13	Co-publications with industrial partners
14	Patents awarded (size-normalised)
15	Industry co-patents
16	Spin-offs
17	Publications cited in patents
18	Income from continuous professional develop.
19	Foreign language bachelor programmes
20	Foreign language master programmes
21	Student mobility
22	International academic staff
23	International doctorate degrees
24	International joint publications
25	Bachelor graduates working in the region
26	Student internships in the region
27	Regional joint publications
28	Income from regional sources
29	Master graduates working in the region

The sunburst seen on Figure 2 can be considered an evolution of the visualisation approach previously adopted by the U-Map project on the European Classification of Higher Education Institutions. Federkeil, Kaiser, et al. (2012) recognise that project as an essential inspiration to U-Multirank, not only because of how results could be presented, but also because it proposed comparing institutions in the face of their missions, profiles, and characteristics.

From a comparison perspective, the sunburst approach becomes an effective tool to visually analyse the profiles of different institutions and see the strengths and weaknesses in the areas of interest of each end-user. This powerful visualisation enables UMR's decision not to produce oversimplified league tables of the world's top universities. Dropping the standard tables also makes it more feasible to go beyond the comparison of internationally oriented research universities to cover profiles such as: regionally oriented colleges, music academies, teacher training colleges, and universities of applied sciences. (Federkeil, File, et al., 2012; van Vught et al., 2012)

In the example of the University of São Paulo, eight indicators are empty, indicating the absence of data for the institution (e.g., graduating on time). The remaining bars are filled in five levels, from "E" (1) to "A" (5). Comparing the list on Figure 2 with Figure 1, we notice some research indicators are not displayed in the chart, such as "professional

publications". Those omissions result from the expansion in the number of indicators since UMR's conception, which were not incorporated into the sunburst after its original design. A consequence seen in the case of USP is that some high-performance results, such as "Open Access publications", are not visible in the chart.

According to Moed (2017), ranking developers have made enormous progress over the past decade, in some cases offering informative, user-friendly systems with series of indicators that allow institutions to be ranked accounting for the diversity of their profiles. That seems to be the case of U-Multirank, which is recognised by some scholars as one of those that better meet the community's expectations of fairness and responsibility, despite existing reservations regarding overall ranking shortcomings (Fanelli, 2016; Gadd et al., 2021; Hazelkorn and Gibson, 2017).

2.2 U-MULTIRANK IN BRAZIL

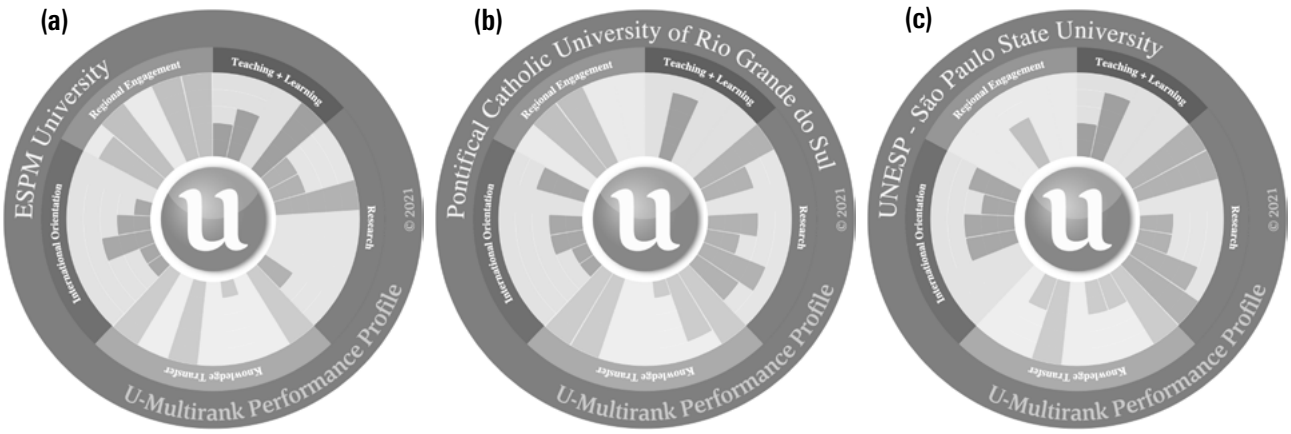
The coverage of Latin American (LA) institutions in U-Multirank is too small to have value within the continent. Of the 1948 HEI currently covered by UMR, only 52 are in LA, 34 in Brazil. The number is far from representative as the Brazilian higher education census reports 2537 HEI active in the country, most of them focused on offering undergraduate degrees (INEP, 2020; U-Multirank, 2021). As detailed in Brasil (2020), a total of 432 of those institutions also offer graduate programs, which may include master's or doctoral courses.

The low representation of Latin America in UMR partly results from the lack of institutional initiatives to register and provide the necessary data for the ranking. The challenge to overcome that problem, however, is made clear by Fanelli (2016, p. 8), who mentions "the quantity and quality of statistics on LA higher education systems vary per country and even per category of institution". The scholar also highlights that only a few Latin American HEI have adequate information about nonresearch indicators available, something evident from the blank indicators in the performance chart of the University of São Paulo (USP), presented in Figure 2.

The lack of complete information from higher education institutions has potentially harmful effects on comparisons, as users may find it challenging to produce their tailor-made lists. For instance, when using U-Multirank's interactive web tool, it is possible to filter HEI based on the subject area, country, and sets of variables associated with the five different dimensions. The resulting list can be sorted alphabetically, based on any particular measure or using "top scores". This system is based on the Olympic medal approach, where the list is ordered according to the number of gold medals won (which would be the "A" scores in UMR), and then by the subsequent levels (U-Multirank, 2021).

Using the described flow to evaluate Brazilian HEI involved in graduate education, in this study, universities were compared as a whole instead of by disciplinary field, and then they were filtered to include those offering master's or doctoral degrees. The selection of indicators was then expanded to include all of the 35 available across the five UMR dimensions. The resulting list of 33 HEI was sorted according to "top scores", and Figure 3 displays the charts for the top three universities.

Figure 3: U-Multirank performance charts for the three top scoring Brazilian HEI offering graduate programs. Interactive version at https://bit.ly/3fRH30m



The best-ranked institutions are shown from Figures 3a to 3c, with the Higher School of Advertising and Marketing (ESPM) as the top-performing HEI in the country. That is a surprising result. While ESPM is a traditional institution with undeniable quality from over 70 years of experience, it is nevertheless a specialised HEI, offering nine undergraduate courses and five graduate programs in communication-related areas. Three of the institution's PPG offer only master's courses, all ranked "regular" by CAPES (grade 3). The other two programs count with doctorates as well and are ranked as "very good" (grade 5). ESPM may be considered a midsize HEI, and it counts with campuses in four different cities (MEC, 2021; CAPES, 2021d; INEP, 2020).

As a direct comparison, the institution shown in Figure 3c – São Paulo State University (UNESP) – counts with 141 PPG (including 139 master's and 116 doctoral courses), 50 of them ranked as "very good", while 27 are considered of excellence (receiving the top grades 6 or 7). At the undergraduate level, UNESP offers 136 courses, 31 assessed as

"excelent", and 104 as "very good" by the Brazilian Ministry of Education. Besides that, the institution has 34 campuses across 24 different cities (MEC, 2021; CAPES, 2021d).

This very superficial comparison tells one main story: ESPM and UNESP are in two different categories, and they should be compared addressing their differences. From an institution list built without purpose, just selecting every possible measure without filtering for subject area, institution size, legal status and other potential indicators, the results become less significant. A proper list should be built with intentionality, exploring institutional profiles from a combination of desired characteristics (e.g., ESPM excels in marketing, communication, and design, but would not even be listed should the user be interested in health sciences).

While analysing the multidimensionality of rankings, Moed (2017, p. 987) concludes that they only allow "looking into the outside world through a few vertical splits in a fence, one at the time. In this sense, these systems are still one-dimensional". A consequence of that perspective is that users also become responsible for the proper use of rankings, as they must decide the best way to look through the fence so they can see what is relevant for them. The problem for U-Multirank in Brazil is that the reduced number of institutions, most with data unavailable for many indicators, leads to few and narrow splits in the fence, making it very hard to see any clear picture on the other side.

2.3 THE PROPOSED MODEL

The concept of a multidimensional assessment for the SNPG matured over the course of more than two years, also counting with the lessons from decades of a robust evaluation system. Many actors and organisations are involved in the process, and views of what the system could and should become are not always uniform. Despite that, the PNPG Committee (2020) presented CAPES' Higher Council with its final proposal for a multidimensional evaluation. The document was unanimously approved and the proposal was given a finality that even contradicted the committee's original expectations (Audy, 2020).

According to Audy (2020, 27:55), the proposal was intended as one of many contributions for the improvement of evaluation in Brazil, as the committee never had the ambition of being in charge of producing a new model by itself, even considering the multiple contributions from the involved organisations. The actual model would come from the work of CAPES and the academic community shaping those inputs. While that might have once been the intention, the idea of a full-fledged evaluation model could not be avoided from its approval by the top instance of CAPES' management, which is its Higher Council.

Audy (2020) also mentioned there were few changes from the previous report by the PNPG Committee (2018), but a significant one was the adjustment of the five original dimensions to fit those adopted by U-Multirank, despite minor nomenclature variations. Now, the new model would consider: Personnel training; Research; Innovation and knowledge transfer; Societal impact; Internationalisation. In complement to that, some of the core suggestions from the PNPG Committee (2020) are listed in summary below:

i) **Grading system** – The result of the evaluation will no longer consist of a single grade for a graduate program. Each of the five dimensions will be graded separately on the already discussed scale from one to seven.
ii) **Accreditation** – In the existing single-grade model, the minimum grade required to renew a PPG accreditation has been three (considering the 1–7 scale). Requirements are yet to be defined in the new system, but the committee suggests three to remain the lowest possible grade for what it considers core dimensions: "personnel training" and "research".
iii) **Indicators** – The proposal includes a series of indicators as suggestions for the assessment of each dimension. While some of them would be new to the Brazilian evaluation, especially those regarding "innovation and knowledge transfer" and "societal impact", most are well established from previous cycles. A major suggestion is that indicators should be universal to all disciplines, and custom ones would not be allowed.
iv) **Funding** – The new model should be taken into account in funding strategies for research and graduate education, as it is suitable for diversification. Regardless, it should not be the only guidance in the decision-making process.
v) **Self-assessment** – An institutional strategic plan should be a fundamental requirement in the evaluation process, serving as a reference for a self-assessment process within the PPG. That should be an essential component for evaluating each dimension of the new model.

While the PNPG Committee (2020) includes other suggestions for the new evaluation model, they are not pivotal to the multidimensional proposal.

3 ASSESSING THE PROPOSAL

The proposal for a multidimensional evaluation of graduate education in Brazil considers that "several recommendations constitute important paradigm shifts and require time for implementation" (PNPG Committee, 2020, p. 27). Because of that, the proposed changes were to be implemented only for the following cycle (2021–2024). However, despite that ambitious statement, one of the findings from this study is that the new model does not change much from the evaluation already in place, and it wastes the opportunity to promote an actual multidimensional assessment.

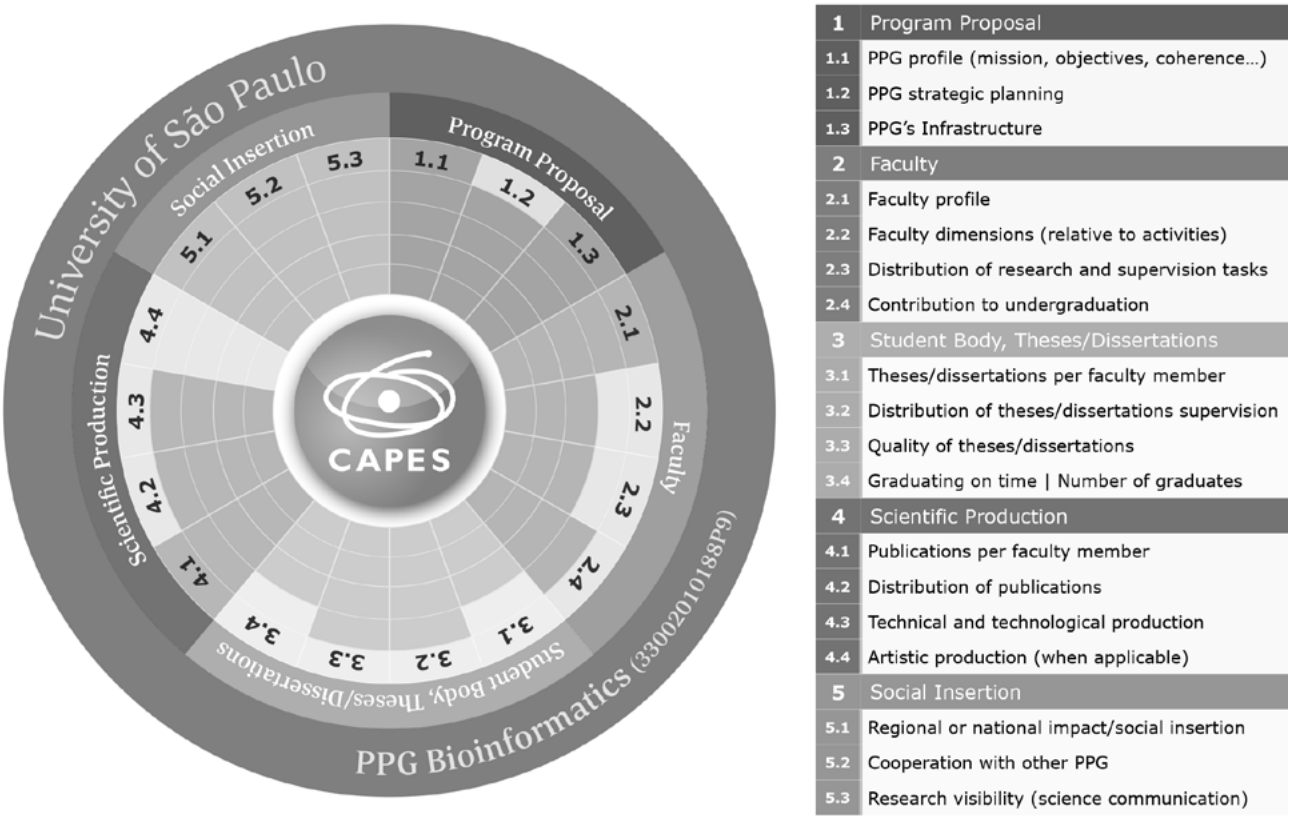
3.1 EVALUATION WAS ALREADY MULTIDIMENSIONAL

One of the essential principles behind Brazilian evaluation is that it should be comparative, so the SNPG can have a transversal equivalence among graduate programs from different disciplines. Thus, a PPG in mathematics is expected to present the same level of quality as one in sociology, provided they have the same grade and respecting inherent characteristics of each area (CAPES, 2010).

To make that possible, CAPES standardised its assessment form in 1998. Each discipline could adapt the proposed indicators to their reality, but they should assess the same set of items: seven in the first version. That number was reduced to five in a subsequent revision (2005–2007), and the form went through additional adjustments with every cycle. The version adopted at the 2017 quadrennial evaluation consisted of two levels: 18 subitems organised into five items. Each subitem is graded on a five-level scale: "insufficient", "weak", "regular", "good", and "very good". (CAPES, 2010; Monteiro et al., 2019).

Despite some terminology differences, the similarity to U-Multirank's methodology is quite apparent. To demonstrate that, Figure 4 shows how the UMR sunburst could be applied to visualise the 2017 evaluation results of the graduate program in Bioinformatics from the University of São Paulo (USP).

Figure 4: Results from CAPES' 2017 evaluation of the PPG on Bioinformatics (USP), transposed to U-Multirank's sunburst



1	**Program Proposal**
1.1	PPG profile (mission, objectives, coherence...)
1.2	PPG strategic planning
1.3	PPG's Infrastructure
2	**Faculty**
2.1	Faculty profile
2.2	Faculty dimensions (relative to activities)
2.3	Distribution of research and supervision tasks
2.4	Contribution to undergraduation
3	**Student Body, Theses/Dissertations**
3.1	Theses/dissertations per faculty member
3.2	Distribution of theses/dissertations supervision
3.3	Quality of theses/dissertations
3.4	Graduating on time | Number of graduates
4	**Scientific Production**
4.1	Publications per faculty member
4.2	Distribution of publications
4.3	Technical and technological production
4.4	Artistic production (when applicable)
5	**Social Insertion**
5.1	Regional or national impact/social insertion
5.2	Cooperation with other PPG
5.3	Research visibility (science communication)

The Bioinformatics program shown in Figure 4 was selected at random from PPG graded four in the 2017 evaluation. The grade was chosen because a multidimensional profile would be easier to visualise for a program with an intermediate level of quality than for one that has weaknesses in all dimensions or excels at everything. The PPG is part of the "Biological Sciences I" evaluation area in CAPES' classification system and counts with a master's and a doctorate in genetics. Its evaluation report is publicly available at CAPES (2021b).

As displayed in Figure 4, the PPG had seven subitems evaluated as "very good" (e.g., 1.1), six as "good" (e.g., 1.2), four as "regular" (e.g., 2.2), and one was not aplicable (4.4, regarding artistic production). The sunburst shows the distribution of strenghts across the five dimensions, and the evaluation report reflects that perception in the aggregation of results, attributing "very good" to dimensions 1 and 5, and "good" to the remaining three. After weighting the five-dimensional results, the PPG received grade four as its final result.

Aggregation and weighting of the assessment items, while transparent through the public regulation of evaluation, have been a major problem in the process. For instance, Marques et al. (2020) performed a statistical analysis to map the probability for each of the five items to influence grade changes in the PPG assessed in 2017. The conclusion was that "scientific production" and "student body, theses/dissertations" had the most impact to achieve better grades, while "social insertion" was of no relevance across all 49 disciplines in CAPES evaluation. Regulations stated that item should count for at least 10% of the PPG grade, and every committee kept that at the minimum so that programs like the one in Bioinformatics could not benefit from the top performance it displayed in that dimension.

The objective of this study is not to discuss the quality of the assessment form that was used during the quadrennial evaluation of 2017, especially since that has already been revised by Monteiro et al. (2019) for the evaluation planned for 2021. The goal here is merely to show that the evaluation process in Brazil has already been multidimensional, even though the results are not.

3.2 MISSED OPPORTUNITIES

Understanding how the evaluation has been organised around an assessment form reveals that the new multidimensional model changes very little in the process. The only real difference is that the results will come from one step before the usual final grade calculation. Thus, considering the inspiration from U-Multirank and its message against composite scores and nontransparent aggregation, opportunities seem to be wasted in the Brazilian proposal.

As it happens in UMR, end-users should be able to select the indicators (or subitems) that would help them understand the profiles of the graduate programs according to their interests. While this user-driven flexibility could increase the complexity of the evaluation process, it would nevertheless produce richer results. Besides that, the proposed rigidity of indicator selection imposed for all disciplines takes even more multi from the multidimensional.

When Moed (2017, p. 987) analysed five of the most prominent rankings of today, he concluded that "a system should not merely present a series of separate rankings in parallel but rather a dataset and tools to observe patterns in multifaceted data". Without that, a national evaluation system that should strive to go beyond what university rankings can offer, ends up behind what U-Multirank already does.

While the Brazilian multidimensional evaluation should go beyond what has been proposed, it is crucial to recognise the multiple grade system as an advancement. The unique composite scores that aimed to define the quality of a graduate program were too outdated, and in previous evaluations a PPG with top performance in social insertion would not be valued as much as another with significant scientific production, as the weight applied to the items in the final grade was unbalanced: usually of 10% for the first and 35% for the later (CAPES, 2017). Why should a PPG's primary mission always be expanding the frontiers of knowledge and never focus on regional or societal impact? A five-grade system is a modest but relevant step to allow graduate programs to find their own identities.

Regardless, the proposed model overlooks another significant opportunity: the lack of attention to self-evaluation. In its 28 pages, the proposal by the PNPG Committee (2020) dedicated only a single paragraph to the issue, while it could be the most powerful instrument in a genuinely multidimensional assessment.

4 SELF-EVALUATION

As stated early in the paper, a self-evaluation strategy has been an integral part of the intended evolution of the Brazilian assessment of research and graduate education. When multidimensionality became a part of that, it was clear that the two initiatives should walk together, but this has not been the case so far.

According to Trevisol and Brasil (2020) there is little literature investigating self-evaluation from the perspective of the SNPG, and the system had almost no experience with those practices. Despite higher education institutions regular development of Institutional Development Plans (PDI), that knowledge was rarely applied in the planning and monitoring of PPG. Thus, while the working group created at CAPES to propose a self-assessment methodology for graduate programs faced a challenge, it also had the opportunity to build something new.

Through Verhine et al. (2019), the working group reported its findings on self-evaluation, proposing strategies for its adoption in the assessment of PPG. The central concept was that each graduate program would implement a custommade process capable of capturing relevant aspects of its mission and objectives, including societal impact, international profile, and distinct scientific decisions. Furthermore, the proposed Brazilian self-evaluation came from a grounded understanding of the SNPG and international inspiration. One of the highlighted countries was the Netherlands, where the group recognised self-evaluation as a unique process, as it is the core of the national assessment and a pivotal instrument leading to the improvement of the country's research units.

The model currently in use to assess the quality of research in Dutch universities is based on six-year cycles, and it is known as Strategy Evaluation Protocol (SEP). An essential lesson from such a model is in its collaborative design since SEP is a joint effort by the Association of Universities in the Netherlands (VSNU), the Netherlands Organisation for Scientific Research (NWO), and the Royal Netherlands Academy of Arts and Sciences (KNAW). With a focus on three dimensions (viability, research quality, and societal relevance), a self-evaluation report is prepared by research units in light of their mission and strategies. Reports can include appropriate indicators to support the presented narrative, but no uniform measure of success is prescribed. That means each unit can choose the best metrics that serve as evidence of its performance, provided they keep away from indicators such as the Journal Impact Factor (not allowed) and the h-index (strongly discouraged). An external assessment committee is then appointed to analyse the self-evaluation document and, after a site visit, an assessment report is produced (VSNU et al., 2020).

Another example mentioned by Verhine et al. (2019) comes from Finland, where a *benchlearning* system was implemented. In it, developing research units could seek established ones for active interaction and commitment to mutual development (Leite et al., 2020). Such a strategy would be very beneficial in an asymmetric country like Brazil, where distinct levels of scientific maturity are observed across PPG. Because of that, it would directly align with the working group's proposal for a formative self-evaluation where complementary site visits could serve as the external assessment element but also as a means of support by more experienced PPG to developing ones (Verhine et al., 2019).

While Verhine et al. (2019) brought additional inspiration and presented a strong proposal for self-evaluation, the current implementation follows the path of the multidimensional model, with missed opportunities. For the current assessment cycle (2017–2020), self-evaluation has been relegated to adding two subitems to the new assessment form proposed by Monteiro et al. (2019). Each subitem – "strategic planning" and "self-evaluation" – would have a recommended minimal weight of 10% only, putting the effort in danger of becoming statistically irrelevant to the final result, as Marques et al. (2020) measured to be the case for "social insertion" in the previous assessment form.

Verhine (2020) recognises that the initial implementation of self-evaluation is very modest. The working group coordinator states CAPES' Higher Education Council (CTC-ES) believed institutions should have time to adapt and build internal assessment infrastructures before the changes significantly impact their grades. The benefit of the initiative right now is in changing the institutional mindset so that a broader self-evaluation model could be implemented in the future. That seems to be a sound decision for the 2021 evaluation. The problem is that the allegedly comprehensive multidimensional model is already planned for 2025, so why is the next level self-assessment not a part of that?

5 CONCLUSION

Leite et al. (2020) see evaluation as a political act that interferes in the life and existence of the evaluated, whether they are people, courses, institutions. CAPES and the evaluation community are aware of that responsibility, so they have strived to evolve assessment strategies continuously, sometimes through minor adjustments to match the dynamic nature of the Brazilian National System of Graduate Education, but other times as fundamental paradigm shifts (Brasil, 2020; CAPES, 2010).

Considering the dimensions of the SNPG and the ever-increasing effort to assess its performance and impact, a central evaluation system was becoming impossible to maintain, and it was progressively unable to capture the complexity of Brazilian science. Therefore, it was time for those paradigm shifts regarding the established one-dimensional perception of quality and the role the evaluated could play in their own evaluation. Hence, a multidimensional approach to quality was required, as was outsourcing part of the assessment task to higher education institutions and graduate programs (PNPG Committee, 2018; Faljoni-Alario et al., 2018; FOPROP, 2018; Verhine et al., 2019).

From the analysis of documents, reports, guidelines, and legislation regulating Brazilian evaluation, this study concludes that the multidimensional model proposed fails to meet expectations. Despite being inspired in part by U-Multirank, the model is based on universal indicators for all disciplines and continuing weighting and aggregation of results (leading to a grade for each of the five suggested dimensions). Furthermore, while improving the items and subitems of the assessment form to reflect the new dimensions, the new model changes very little in the evaluation process. As this study shows, any multidimensionality proposed was already a part of the system, only lost through weighting and aggregation. Those have improved but will still be there, hindering future results as well.

While the multidimensional proposal took a step towards addressing one of the longest standing flaws in the Brazilian evaluation – which was applying a broad benchmarking strategy to a heterogeneous system – it could have benefited from further inspiration. For instance, from U-Multirank, it could have learned to drop the aggregated grading system altogether. Likewise, inspiration from models such as the Dutch Strategy Evaluation Protocol could lead to a revision in the fixed indicator perspective, as each PPG could have the freedom to choose the ones that could better serve as evidence of their accomplishments. However, that would be an impossible strategy unless self-evaluation becomes part of the process, as it was initially intended.

The self-evaluation proposed by Verhine et al. (2019) seems to be sound and well structured, and it deserves more protagonism in the overall assessment process. Self-evaluation should not be limited to a couple of underweighted subitems; it should mean freedom for graduate programs to present unhindered narratives of their accomplishments, based on their missions, and supported by evidence that makes sense, instead of indicators that are easy to compare. It should allow PPG to be different, knowing that the quality of their work would not need to be reshaped to fit formulas incapable of measuring diversity. Finally, it should allow multidimensionality to manifest as more than a set of disconnected resulting grades but as a process that recognises and values dimensions in the core of the evaluation process.

Multidimensionality is necessary for a better evaluation in Brazil, and it is only truly possible through self-evaluation. The strategies complement each other and must be articulated in such a way as to heighten their potential individual effects.

ACKNOWLEDGEMENTS

The author would like to thank the following scholars for their inspiring comments and suggestions on the first draft of this paper: Frans Kaiser, Ludo Waltman and Robert Tijssen (from U-Multirank), Robert Verhine and Denise Leite (from the Brazilian self-evaluation effort) and Sarah de Rijcke (from CWTS).
This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

REFERENCES

Audy, J. L. N. [Unesp Oficial]. (2020). O novo modelo de avaliação da CAPES [webinar].
Retrieved August 13, 2021, from https://youtu.be/TyNrLIiInHU

Balbachevsky, E. (2005). A pós-graduação no Brasil: Novos desafios para uma política bem-sucedida. In S. Schwartzman & C. Brock (Eds.), Os desafios da educação no Brasil (pp. 276–304). Nova Fronteira.

Brasil, A. (2020). Building a national system of research and graduate education: How did the university become the house of science in Brazil? Revista Nupem, 12(27), 222–253.

BRASIL. Ministério da Educação. (2021). Cadastro e-MEC de Instituições e Cursos de Educação Superior. https://emec.mec.gov.br

Calero-Medina, C., López-Illescas, C., Visser, M. S. & Moed, H. F. (2008). Important factors when interpreting bibliometric rankings of world universities: an example from oncology. Research Evaluation, 17(1), 71–81.

CAPES. (2010). Plano Nacional de Pós-Graduação - PNPG 2011-2020 (Vol. 1). CAPES.

CAPES. (2017). Portaria nº 59, de 21 de março de 2017, Aprova o regulamento da Avaliação Quadrienal (legislation). Diário Oficial da União. Brasília, DF.

CAPES. (2018). Avaliação Quadrienal em números (Rep.). Brasília, DF. https://bit.ly/2Y2rPvO

CAPES. (2019). Aprimoramento do processo de avaliação da pós-graduação (Rep.). CAPES. Brasília, DF. https://bit.ly/3wDRgDU

CAPES. (2021a). Avaliação Quadrienal. Retrieved June 18, 2021, from https://bit.ly/35wv16S

CAPES. (2021b). Sistema de Informações Georreferenciadas [Geocapes]. Retrieved August 20, 2021, from https://geocapes.capes.gov.br/

Comissão Especial de Acompanhamento do PNPG 2011-2020. (2018). Proposta de aprimoramento do modelo de avaliação da PG (Rep.). CAPES. Brasília, DF. http://shorturl.at/glvxE

Comissão Especial de Acompanhamento do PNPG 2011-2020. (2020). Proposta de aprimoramento da avaliação da pós-graduação brasileira para o quadriênio 2021-2024: Modelo Multidimensional (Rep.). CAPES. Brasília, DF. https://bit.ly/3wyR097

Conselho Federal de Educação. (1965). Parecer CFE nº 977, de 3 de dezembro de 1965. Estabelece a definição dos cursos de pós-graduação no país (legislation). Brasília, DF.

Faljoni-Alario, A., Junior, C. F. d. S., Brito, E. P. Z., Gontijo, J. A. R., Romero, M. A., Santos, P. J. P. d. & Canuto, S. (2018). Avaliação da pós-graduação: Considerações do CTC-ES (Rep.). Brasília, DF.

Fanelli, A. (2016). U-Multirank and Latin American Universities. International Higher Education, (84), 7–9.

Federkeil, G., File, J., Kaiser, F., van Vught, F. & Ziegele, F. (2012). An interactive multidimensional ranking web tool. In F. van Vught & F. Ziegele (Eds.), Multidimensional ranking: The design and development of u-multirank (pp. 167–177).

Federkeil, G., Kaiser, F., van Vught, F. & Westerheijden, D. F. (2012). Background and design. In F. van Vught & F. Ziegele (Eds.), Multidimensional ranking: The design and development of u-multirank (pp. 85–96). Springer.

Ferreira, M. d. M. & Moreira, R. d. L. (Eds.). (2002). CAPES, 50 anos: Depoimentos ao CPDOC/FGV. FGV.

FOPROP. (2018). Avaliação da pós-graduação: Proposições à Comissão Especial do PNPG. Brasília, DF.

Gadd, E., Holmes, R. & Shearer, J. (2021). Developing a Method for Evaluating Global University Rankings. Scholarly Assessment Reports, 3(1), 1–19.

Hazelkorn, E. & Gibson, A. (2017). Global science, national research, and the question of university rankings. Palgrave Communications, 3(1), 21.

INEP. (2020). Censo da Educação Superior 2019. Retrieved August 17, 2021, from https://bit.ly/3ABVsFK

Leite, D., Verhine, R., Dantas, L. M. V. & Bertolin, J. C. G. (2020). A autoavaliação na Pós-Graduação (PG) como componente do processo avaliativo CAPES. Avaliação: Revista da Avaliação da Educação Superior (Campinas), 25(2), 339–353.

Marques, C. A., Veiga, A. d. C. & Borges, L. M. C. (2020). A avaliação da pós-graduação no Brasil: resultados e determinantes da avaliação da Capes (2013-2016). Revista Meta: Avaliação, 12(37), 979–1003.

Marques, F. (2019). Avaliação em cinco dimensões. Revista FAPESP, (286), 28–31. Retrieved June 16, 2021, from https://bit.ly/3iNeIKW
Martins, C. B. (2018). As origens da pós-graduação nacional (1960-1980). Revista Brasileira de Sociologia, 6(13), 9–26.

Moed, H. F. (2017). A critical comparative analysis of five world university rankings. Scientometrics, 110(2), 967–990.

Monteiro, A. L., Brasil, A., Horta, B. L., Mamiya, E. N., Federizzi, L. C., Verhine, R. E. & Júnior, W. R. d. S. (2019). Ficha de Avaliação (Rep.). Brasília, DF. https://bit.ly/2PmAf0q

Trevisol, J. V. & Brasil, A. (2020). A autoavaliação nas políticas da pós-graduação brasileira. Anais da XIII Reunião Científica da ANPEd-Sul.
U-Multirank. (2021). U-Multirank: Universities compared your way. Retrieved August 9, 2021, from https://www.umultirank.org

van Raan, A. F. J. (2005). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. Scientometrics, 62(1), 133–143.

van Vught, F., Westerheijden, D. F. & Ziegele, F. (2012). Towards a new ranking approach in higher education and research. In F. van Vught & F. Ziegele (Eds.), Multidimensional ranking: The design and development of u-multirank (pp. 1–7). Springer.

Verhine, R. E. [Unesp Oficial]. (2020). Autoavaliação dos PPG e planejamento estratégico institucional da pós-graduação [webinar]. Retrieved August 20, 2021, from https://youtu.be/GNdYu18KRts

Verhine, R. E., Lima, A. C., Carvalho, D. B. B., Leite, D., Bertolin, J. C., Dantas, L. V., Oliveira, N., Oliveira, R., Avellar, S. & Rosa, D. L. (2019). Autoavaliação de Programas de Pós-Graduação (Rep.). Brasília, DF. https://bit.ly/3jTSRS8

VSNU, KNAW & NWO. (2020). Strategy Evaluation Protocol (2021–2027) (Rep.). The Hague. https://bit.ly/3wAFbzi

Waaijer, C. J. (2018). Why the public nature of university rankings torpedoes their value for benchmarking: an institutional researcher's opinion. Proceedings of the 23rd International Conference on Science and Technology Indicators, 1031–1037.

AUTHOR

ANDRÉ BRASIL
Centre for Science and Technology Studies (CWTS), Leiden University
Willem Einthoven Building, Kolffpad 1
2333 BN - Leiden, The Netherlands
E: a.l.brasil@cwts.leidenuniv.nl
ORCID: 0000-0003-1832-5199

KEYWORDS: policy analysis, evidence-informed policy, research evaluation, multidimensional evaluation, self-evaluation.

CO-DEVELOPING AN IMPACT MODEL FOR EVALUATING THE SOCIETAL IMPACT OF PARTICIPATORY RESEARCH

MAGDALENA WAILZER AND LAURA SOYER

DOI: 10.22163/fteval.2022.547

ABSTRACT

In recent years, an increased focus on societal impact of research unfolding through productive interactions between stakeholders and participatory research processes has been seen. These complex interventions call for more flexible and participatory evaluation processes. This paper sets out to describe the co-creative development of an Impact Model and Reflection Instruments by different stakeholders that make desired and expected societal effects of participatory research visible, and enable a systematic evaluation of these expected changes.

The aim of the Impact Model and the (modular) set of Impact Reflection Instruments is first and foremost to support researchers in the planning and evaluation of societal impacts of their participatory research approaches. In addition, we share the design of the co-development phase and reflections that serve as practical guidance for evaluators who aim to apply theory-based models in participatory settings in other contexts. Finally, the Impact Model and Reflection Instruments aim to enable increased comparability across research projects with participatory research approaches.

I INTRODUCTION

Societal impact unfolds through creative interactions, relationships, and dialogue with external stakeholders (Spaapen et al. 2011, van den Akker & Spaapen 2017). Participatory approaches – the active involvement and engagement of stakeholders in research (Cargo & Mercer 2008) – can therefore serve as an important lever to increase the societal impact of research.

But how can the societal impact of participatory research be evaluated? Following a literature overview on existing evaluation approaches in the field of societal impact and participatory research, this paper presents the Impact Model and Reflection Instruments, which have been co-created bottom-up together with a diverse set of stakeholders.[1] The Impact Model supports the planning of societal impact of participatory research activities. The Impact Reflection Instruments comprise a toolkit of evaluation instruments to assess societal impacts of participatory research.[2] We further shed light on practical experiences gained in structuring and managing the bottom-up approach. As such, we provide a contribution to the description of advantages, challenges and possible solutions of this process, as well as requirements to ensure successful results. In doing so, it is intended to support others who want to implement a similar process.

II BACKGROUND AND LITERATURE REVIEW[3]

II.1 DEFINITIONS AND CONCEPTS

II.1.1 WHAT IS SOCIETAL IMPACT?

Societal impact is often defined as "the demonstrable contribution that excellent research makes to society and economy" (UKRI website 2021). It occurs through creating and sharing new knowledge and innovation, inventing new products, companies and jobs, developing new and improving existing public services and policy or enhancing quality of life and health (Smith et al. 2020). In contrast to scientific impact (i.e. impact that is generated by research within the scientific and disciplinary community itself), societal impact of research focuses on the effects and changes that research activities unfold beyond academia in other areas of life such as society, culture, public services, health or the environment. This can include changes in practice, policy and legislation, as well as changes at the level of awareness, understanding and personal skill development (Smith et al. 2020). Definitions and types of societal impact, as well as understandings of "positive impact", vary according to disciplinary traditions and fields (Reed 2016, Oancea 2013).

1 See: https://ois.lbg.ac.at/en/projects/impact > "Tools" > "The OIS Impact Model"
2 See: https://ois.lbg.ac.at/en/projects/impact > "Tools" > "OIS Impact Reflection Instruments"
3 For further details see: https://ois.lbg.ac.at/en/projects/impact > "About" > "Our Mission" > "What is Societal Research Impact. A Short Literature Review"

II.1.2 WHAT IS PARTICIPATORY RESEARCH?

Participatory research refers to actively involving and engaging stakeholders in the research process through various research designs, methods and frameworks (Cargo & Mercer 2008). Rather than subjects of research, stakeholders become part of the research process (Vaughn & Jacquez 2020); research is not carried out 'to', 'about' or 'for', but 'with' or 'by' them (Hayes et al. 2012). Stakeholders can participate in various stages along the research cycle (Hoekstra et al. 2020) and hold various degrees of power (Arnstein 1969).

Participatory approaches in science and research share overlaps with related concepts from different research fields and contexts, including:

- Open Innovation in Science (OIS), an umbrella term which refers to 'opening up' the scientific process through various strategies, such as applying open innovation approaches from business and industry (Beck et al. 2020). OIS is often also referred to in the context of Open Science and Open Data in Science, where the focus is specifically on the free use, re-use, distribution and publishing of scientific knowledge without legal, technological or social restrictions.
- Citizen Science, originally coined as a method to generate large amounts of data (Bonney et al. 2009), now applied more generally as an intentional engagement of the public in scientific research (Philips et al. 2018)
- Participatory Action Research, where researchers and practitioners collaborate to enable action (Baum et al. 2006)
- Transdisciplinary Research, describing efforts by researchers from different disciplines as well as external stakeholders, working jointly to create new conceptual, theoretical, methodological innovations that integrate and move beyond sectorial and discipline-specific approaches to address a common problem (Klein 2013)
- Responsible Research and Innovation (RRI), widely used in the context of the European Union, which aims to foster inclusive and sustainable research and innovation through co-creation and co-production with society (Owen & Pansera 2019)
- Patient Engagement or Public and Patient Involvement, which, although lacking a common definition, refers to the active involvement of patients in health care (Gallivan et al. 2012)

II.1.3 THE CONTRIBUTION OF PARTICIPATORY RESEARCH TO SOCIETAL IMPACT

For scientific expertise to move into practice and policy settings, where it can progress towards societal impact, it needs to be disseminated and mobilized (Phipps et al. 2016). Only when research evidence, tools and methodologies are used to inform policy or practice outside a purely academic setting, it can unfold societal impact and have an effect on the lives of beneficiaries. This concept is referred to as *Research Uptake* (Phipps et al. 2016). Implementation and practical use of research findings are thus dependent on non-academic partners and stakeholders applying and using them.

Societal impact unfolds in a non-linear way along different pathways – the *Pathways to Impact*. The process of creation is thus inherent to the societal impact produced. In order to achieve research uptake and thus ultimately achieve societal impact, *productive interactions* between researchers and non-academic actors along the research process are vital. Productive interactions refer to exchanges between researchers and stakeholders where both scientifically robust and societally relevant knowledge is produced and valued (Spaapen et al. 2011). Whenever these interactions lead to efforts by non-academic actors to use or apply research results – research is taken up into practice – they are productive. This means, productive interactions represent moments where societal stakeholders influence scientific actors and vice versa (Muhonen et al. 2019) and where both scientific and societal value is generated as a result.

The concepts of research uptake and productive interactions show that societal impact is less about a specific outcome or end product, but more about a process of relationship-building, dialogue and engagement with different research audiences throughout the research process. Therefore, co-production and collaboration between researchers and other stakeholders along the impact pathway usually accelerates the creation of societal impact of research (Phipps et al. 2016). Participatory research approaches, marked by interactive processes that aim to generate knowledge collaboratively through trust, dialogue and collaborative partnerships, can thus serve as important vehicles to achieve societal impact (Greenhalgh et al. 2016, Reed 2016).

II.2 EVALUATING SOCIETAL IMPACT AND PARTICIPATORY RESEARCH

II.2.1 EVALUATING SOCIETAL IMPACT

In recent years, there has been an increasing shift of focus on the contributions and value of science for society - a focus on societal impact of research (Bornmann 2012). However, societal impact of research proves more difficult to be attributed and evaluated than scientific impact (Smith et al. 2020). Challenges include causality and attribution due to complex environments and simultaneous developments, as well as long time spans for societal impact to unfold (Felt & Fochler 2018). Basing impact evaluation on simplistic, linear assumed relationships between research evidence and positive societal change is thus unwise (Felt & Fochler 2018, Smith et al. 2020, Rymer 2011). Evaluation approaches thus have to be considered carefully, as they can lead to unintended consequences of incentivizing, measuring and rewarding impact (Smith et al. 2020).

In the context of societal impact evaluation, evaluation methods are best understood as the process of collecting, contextualizing and interpreting data to assess significance, reach and attribution of societal impacts from research (Reed et al. 2021, Bornmann 2012). They include quantitative measurements, qualitative approaches in form of narrative accounts and case studies and approaches that emphasize interaction, communication patterns and knowledge mobilization between research and societal stakeholders.

Following a typology provided by Reed et al. (2021), three major evaluation approaches of societal impact are known:

- Systems analysis methods are usually used ex-post to examine whether a particular research activity or project was necessary to cause or make a significant contribution to societal impact. They combine a range of qualitative (e.g. interviews, questionnaires, focus groups) and quantitative (e.g. process-based mod-

els such as modelling techniques) research methods to allow a detailed understanding and mapping of causal links from research to impacts (Reed et al. 2021).

- *Indicator-based approaches* are often used for societal impact planning at the beginning of or before the start of a research activity (Reed et al. 2021). At the heart of these are *Theory of Change approaches* (e.g. logic models), linking resources and activities to outputs, outcomes and impact through causal chains and equipped with indicators (Reed et al. 2021) (see chapter 2.4). One prominent example is the Payback Framework mostly used in a health service research context (Donovan & Hanney 2011, Bornmann 2012).
- *Textual, oral and arts-based methods* build impact narratives and cases detailing in how far research activities were necessary to cause societal impact using multiple sources of evidence for attribution (Reed et al. 2021). Textual methods (e.g. qualitative data from interviews and focus groups) enable a more nuanced, subjective understanding of lived experience and values (Reed et al. 2021). Arts-based methods (e.g. participant observation, oral history and storytelling, as well as poetry, fiction, dance, theatre) are especially fruitful in situations where access to the emotional realms of life is desirable or when working with vulnerable groups (Reed et al. 2021).

II.2.2 EVALUATING PARTICIPATORY RESEARCH

With an increase in participatory research practices, evaluating such endeavors becomes increasingly important to justify the resources invested and show proof of the multitude of potential they promise (Barber et al. 2012, Kieslinger et al. 2018, Reed et al. 2018). However, due to the non-linearity and complexity of participatory research processes and pathways, evaluation is challenging or even not feasible (Reed et al. 2018, Barber et al. 2012).

The evaluation of participatory research can focus on the design and process or on the outcome and impact of participatory research (e.g. Kieslinger et al. 2018, Boivin et al. 2018, Reed et al. 2018). Impact can be separated into different domains: In a recent screening of relevant literature in the field, Wehn et al. (2021) distinguish between societal impact, economic impact, environmental impact, science and technology impact and governmental impact. Most evaluation frameworks aim to assess the field of societal impact where differentiations are made between the impact on individuals and collective impact. Drawing from practice, Barber et al. (2021) find that evaluating the impact of participatory research on individuals (e.g. researchers, members of the public) is more feasible than evaluating the overall quality, usefulness or large-scale impact of participatory research. Individual-level societal impact indicators could include indications of new knowledge or skills, challenged assumptions or deepened understanding.

While the evaluation of design and process of participatory research is not the main focus of this paper, it should be noted that evaluating the design of participatory research processes early on could allow for an adaptation, which in turn improves the delivery or outputs of them and thus increases the likelihood of impacts arising from it (Reed et al. 2018). In practice, participatory processes are evaluated more often, while outcomes, if reported, are mainly self-reported and perceived (Boivin et al. 2018, Bührer et al. 2021).

Literature largely calls for a mixed methods approach to evaluate participatory research (e.g. Kieslinger et al. 2018, Barber et al. 2012, Wehn et al. 2021). Evaluation of participatory research could thus consist of a combination between different methods such as observations, qualitative interviews, quantitative surveys, statistics, focus groups and document analyses (Kieslinger et al. 2018, Wehn et al. 2021).

Reed et al. (2018) suggest that some methodologies common in societal impact evaluation could also serve to evaluate participatory research. For example, *Theory of Change approaches* (e.g. logic models) could be used by structuring them around the goal pursued with the engagement. *Contribution analyses* could provide an account of the contribution story of participatory research for each stage of the pathway. Finally, *outcome mapping* could help to identify the changes desired in participating partners, develop strategies to achieve them and then monitor these changes to track them (Reed et al. 2018).

There are no commonly established indicators to evaluate participatory research (Kieslinger et al. 2018) and specific indicators provided in participatory research evaluation frameworks are scarce (Wehn et al. 2021).

II.2.3 PARTICIPATORY EVALUATION

Participatory evaluation refers to situations where stakeholders of a program or policy are involved in evaluation decision-making and reports together with an external evaluator (Turnbull 1999). Thus, participatory evaluation is conducted through a partnership between professional evaluators, as well as practitioners, policy decision-makers or the interested public of a program or policy (Cousin & Whitmore 1992). Participatory research approaches call for participatory evaluation. Bornmann (2012) argues that qualitative evaluation of societal impact should not be dominated exclusively by scientists, as they often have trouble discerning the societal impact of research. Instead, researchers and external stakeholders should jointly evaluate participatory research.

On the one hand, participatory evaluation increases the likelihood of the use of evaluation results and provides a voice to often unheard groups, which enriches public debate. However, poor quality evaluation through unbalanced participation or unresolved conflict decreases the likeliness of results to be used (Plottu & Plottu 2011). While evidence of benefits of participatory evaluation on learning, evaluation capacity building or the use of evaluation can be found, contradictory or unintended effects are also documented (Smits & Champagne 2008). For participatory evaluation to be effective, the process needs to be properly managed so as to ensure a balanced expression of viewpoints among participants (Plottu & Plottu 2011).

II.2.4 THEORY OF CHANGE APPROACH

To respond to the complex nature of evaluating the societal impact of participatory research, more flexible evaluative approaches, such as developmental (e.g. Patton 1994) or realist (e.g. Pawson & Tilley 1997) evaluations are required. Theory-based evaluations, which identify and test causal processes, are particularly suitable to evaluate research contributions in complex systems and thus provide potential to evaluate participatory research and research that crosses disciplinary boundaries (Belcher et al. 2020). Recommended and widely used in practice in both

societal impact evaluation and participatory research evaluation are Theory of Change approaches (Reed et al. 2018, Reed et al. 2021).

Theory of Change approaches show how different resources, inputs and activities are linked to specific outputs, outcomes and overall impact and objective. They trace causal chains from research to impact based on anticipated logical frameworks or a Theory of Change. They involve the identification of activities, impact indicators and research objectives, either through an expert-led top-down or co-creative bottom-up approach with relevant stakeholders. One advantage is their ability to standardize the collection of data in the creation of case studies that are easily comparable and thus transferrable to different disciplinary contexts (Reed et al. 2021).

Theory of Change approaches can be used as an integrative framework for the design and analysis of evaluations using multiple methods. Data from multiple methods can be analyzed and interpreted together in order to enable greater insight into a program's operations and effectiveness (Caracelli & Greene 1997).

III THE IMPACT MODEL AND REFLECTION INSTRUMENTS

This chapter explores the intention and goal as well as the approach used – based on the reviewed literature – to develop an underlying model for evaluating societal impacts. It further details the development process of the Impact Model and the Reflection Instruments.

III.1 INTENTION AND GOAL

The Ludwig Boltzmann Gesellschaft Open Innovation in Science Center (LBG OIS Center) is a Competence Center and a leading international hub for investigating and experimenting with Open Innovation research practices. Since its foundation in 2016, the LBG OIS Center has been supporting and enabling researchers in applying participatory research. In the course of our work, we wanted to know: (a) How do we know that participatory research leads to societal impact and, most importantly (b) what changes does participatory research lead to. Rather than evaluating the overall societal impact of a specific participatory research project, we wanted to find a mechanism to evaluate effects of the participatory approach used in a research project. This would allow us to draw comparison on the societal impact of participatory research projects (due to the participatory methods applied) across all fields and disciplines.

Therefore, we set out to co-creatively develop a comprehensive Impact Model for participatory research that would describe how participatory research leads to societal changes and, ultimately, societal impact. In addition, we wanted to develop an evaluation toolset (that we called Reflection Instruments) to systematically evaluate these societal changes. The resulting Impact Model should help researchers to plan their participatory research projects and to identify societal impacts. The Reflection Instruments, meanwhile, should support them in establishing simple and practicable mechanisms to receive regular feedback on whether and in how far these societal changes are being achieved through their participatory research.

III.2 ANCHORAGE IN LITERATURE

To develop the Impact Model, we focused on societal impacts, rather than scientific impacts. In line with the recommendation formed by Barber et al. (2012), we focused on societal impacts on an individual scale. As previously suggested (see chapter 1.3), we adhere to societal impact concepts suggesting that collaborative partnerships will lead to an increased relevance of research findings and an increased motivation to use and apply these findings in (institutional) practices. Therefore, we consider the pathway of societal impact through individuals a valid instrument to achieve research uptake and thus societal impact on a collective and institutional level. At the same time, the perspective of societal impact through the lenses of individuals allows for a more direct and practical implementation of the Reflection Instruments.

As Theory of Change approaches are common in both the evaluation of societal impact and of participatory research (Reed et al. 2018, Reed et al. 2021), we decided to follow the structure in the development process of the Impact Model. For the development of the Reflection Instruments, we aimed to have a mixed-methods approach represented as recommended in literature (e.g. Wehn et al. 2021, Reed et al. 2021, Staley 2015). While we were aware of the multitude of challenges in evaluating societal impact, such as the time lag between the participatory research conducted and the societal impact unfolded (see chapter 2.1), we wanted to focus on practicability and usability of the Impact Model and Reflection Instruments for researchers. Therefore, we decided to rely on Reed et al.'s (2018) suggestion that the evaluation of the design and process of the participatory research project can ultimately increase the likelihood of impacts arising from them, as interventions throughout the process are made possible. The development of Reflection Instruments was thus aimed to enable a use during the process of a participatory research project. This is where collecting data is most feasible for researchers and the information the data provides serve as important anchors for societal impact.

Many frameworks are developed top-down rather than co-created beyond the piloting phase (e.g. Boivin et al. 2018). While the last step – refining and sharpening the indicators of the Reflection Instruments – was not conducted collaboratively, we tried to go a step further than merely piloting the Impact Model and Reflection Instruments collaboratively and also involved our stakeholders in identifying and defining methods and indicators. This step was particularly new, as non-evaluation experts were invited to develop evaluation tools.

III.3 THE DEVELOPMENT PROCESS

To develop the Impact Model, as well as instruments and methodologies to reflect on the impact elements, we worked together with *Measury*[4], a social research organization, who supported us in the design and implementation of the co-creative process.

4 See: https://www.measury.eu/

III.3.1 DEVELOPING THE IMPACT MODEL

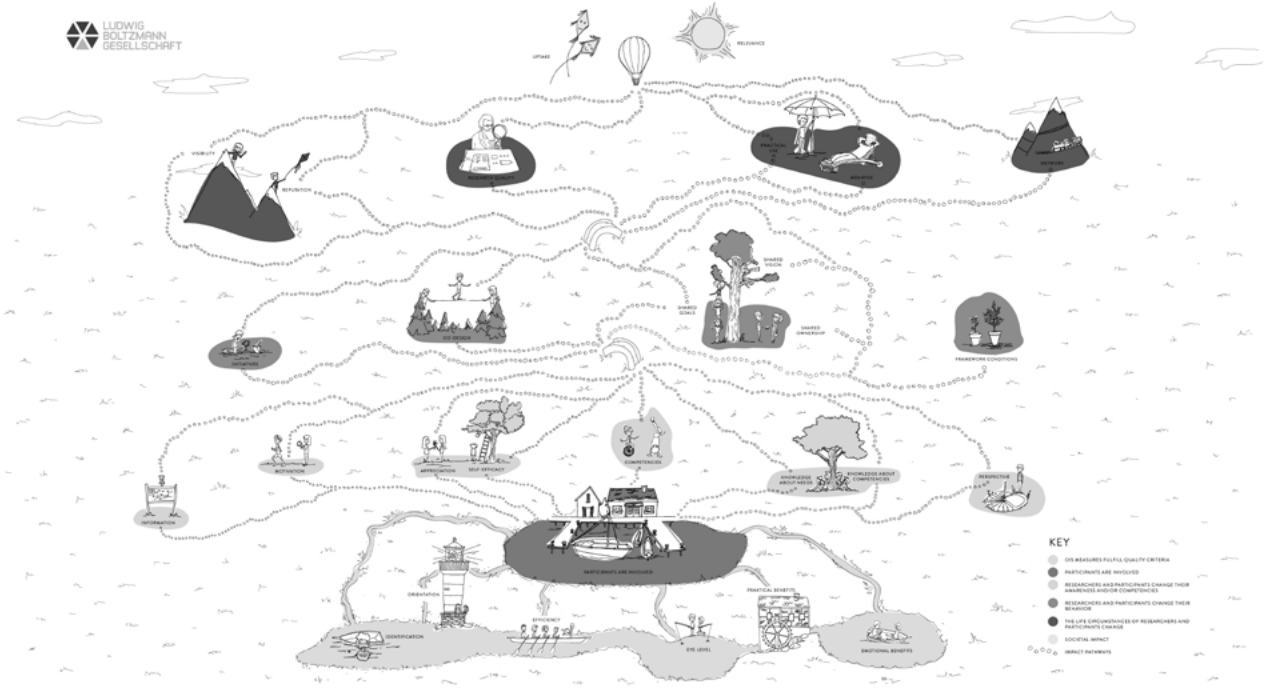
Based on the theoretical concept of the Theory of Change and a stakeholder mapping process, two co-creative workshops were designed around the question "What effects should participatory research approaches have". To reflect on this question, four initiatives of the Ludwig Boltzmann Gesellschaft (the initiative *Crowdsourcing Research Questions in Science,* the interactive workshop design *Ideas Lab*, two capacity building programs, open governance structures[5]) were chosen as examples to guide the process.

The first workshop was held in November 2018 and brought three stakeholder groups from within and outside the LBG to the table: researchers, who work with or are affected by participatory research approaches, practitioners, who are affected by participatory research approaches (e.g. professionals who apply research results in practice) and the community, such as patients or the public affected by participatory research approaches (e.g. by providing expert knowledge). To enable the application process of a complex topic, the activities of the mental health research groups *D.O.T. – The Open Door*[6] and *Village*[7] were used as a basis for application during the workshop. All workshop participants had had previous experience with participatory research in different roles and could therefore speak from personal experience of changes arising due to participation.

The second co-creative workshop in February 2019 brought three stakeholder groups from within and outside the Austrian research context to the table: policy-makers, government actors who formulate policies and thus influence the use of participatory research approaches (e.g. by passing laws), funders, who provide funding and thus influence the use of participatory research approaches (e.g. by defining funding criteria), and the media and thought leaders, non-governmental organizations, journalists or other activists representing civil society at large and thus influence the use of participatory research approaches (e.g. by promoting their value). Due to the macro-perspective of these stakeholder groups, no specific research project was used as a basis for application; rather, participatory research and its long-term effects on the political and societal landscape in general were considered.

In the workshops, participants brainstormed desired changes from participatory research in small groups. Then, they structured and developed pathways together. The effects identified in both workshops were then synthesized and laid out in a unified Impact Model, which underwent several feedback cycles and was continuously refined and sharpened. The Impact Model was then transferred to a playful illustration with symbols for each element, corresponding definitions and affected stakeholders.[8] It encourages a deepened discourse with the different pathways to impact of research involvement methods.

Figure 1 The Impact Model (see footnote 7 for higher resolution picture)



5 See: https://ois.lbg.ac.at/en
6 See: https://dot.lbg.ac.at/
7 See: https://village.lbg.ac.at/
8 See: https://ois.lbg.ac.at/en/projects/impact > "Tools" > "The OIS Impact Model"

The Impact Model consists of different elements: first, it demonstrates what quality criteria participatory research approaches should fulfill (Output / Level 2) to allow effective participation. Once involvement takes place (Output / Level 3), it shows how this involvement will lead to a change in awareness, knowledge, attitude, motivation, skills (Outcome / Level 4), change in actions and behavior (Outcome / Level 5), and, finally, change in the life circumstances (Outcome / Level 6) of all participants involved in a participatory research process. Together, these effects foster more societally relevant research and lead to a higher probability of societal uptake of research results (Impact / Level 7).

To test the Impact Model, we invited the LBG research group Village to apply the model to their research approaches. The research group Village aims to support children of mentally ill parents through building networks of formal and informal support systems in Tyrol. In the co-development phase of the research group, the research team involved practitioners in designing these practice approaches and tools in order to identify support structures for children with mentally ill parents. The application of the Impact Model to practice introduced an additional impact element and allowed to refine existing definitions in the Impact Model.

III.3.2 DEVELOPING THE REFLECTION INSTRUMENTS

To derive useful methodologies and instruments to evaluate and reflect on the impact elements of the Impact Model, we worked together with an ongoing participatory research project: the mental health research group *Village* and participants of their co-development process. In a co-creative workshop with participants of the involvement process in November 2019, we worked with researchers – who set up and were involved in the co-development process of the LBG research group Village and practitioners –, experts of practice (e.g. social workers, psychologists) – who were involved in the co-development process of the research group Village –, as well as community members (i.e. patients and the public). The challenge we faced was to co-develop evaluation approaches together with non-evaluation experts. In order to do this, we asked participants of the workshop to consider each change element individually and ask themselves how they saw this change unfolding in their own experience in the participatory process. After collecting examples from their own experience, they then brainstormed ways and methods how this change could be seen and thus evaluated. In addition, they were specifically asked to raise expectations on the kind of information that is meaningful to them. Then, participants selected a change element and built and pre-tested prototypes of Reflection Instruments.

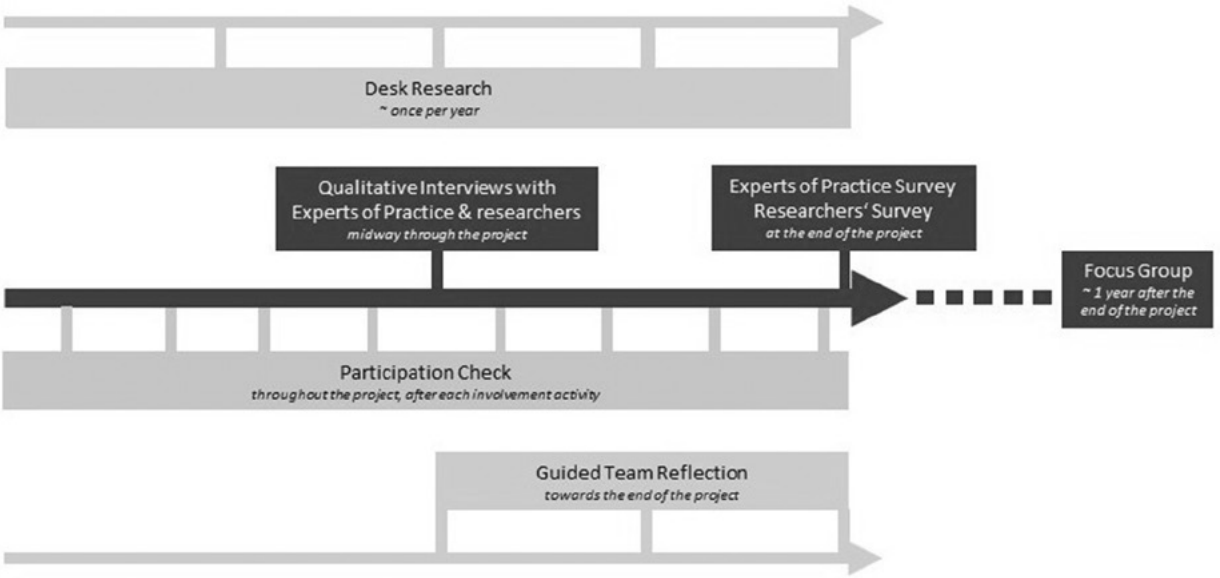
Based on the input received from participants of the participatory research process at the research group Village and in line with the literature (e.g. Reed et al. 2018), we used a mixed-methods-approach to evaluate the different change elements of the Impact Model. These change elements serve as a basis for the Impact Reflection Instruments: a set of different instruments to be used at different time points along a research project applying participatory research approaches.[9] Together, the Reflection Instruments provide a comprehensive picture of the extent to which these methods lead to changes in awareness and competencies, behavior and life circumstances of all stakeholders and, ultimately, societal uptake and impact. They include:

- **Participation Check:** A template for a short questionnaire that provides instant feedback after all participatory research activities (e.g. workshops, meetings) on the design of these participatory research approaches based on quality criteria defined in the Impact Model.
- **Desk Research**: A guideline with suggestions for data to be collected in low-effort desk research (e.g. social media or website analysis, participant lists) as a basis for regular analyses.
- **Qualitative Interviews with Experts of Practice and Researchers:** Two templates for qualitative interviews with both Experts of Practice (i.e. all practitioners, members of the public or policy makers involved in the research project) and researchers to be applied mid-way through the project. These interviews provide valuable insights to adapt and improve the project design.
- **Guided Team Reflection:** A guideline for discussion items for an internal reflection meeting of the research team responsible for designing the research process at later stages of the process. The aim is to reflect on the involvement process and identify late-stage interventions and implications for future research projects.
- **Quantitative Surveys with Experts of Practice and Researchers:** Two questionnaire templates for both Experts of Practice (i.e. all practitioners, members of the public or policy makers involved in the research project) and researchers to be applied at the end of the project. The aim is to provide insights into the experience of the entire involvement process, as well as to uncover short-term effects and impacts.
- **Focus Group:** A guideline and question template for a focus group reflection meeting with a selected group of stakeholders to be applied a year after the project has ended. The aim is to uncover medium- and long-term effects and impacts of the involvement process.

The Reflection Instruments can be used as a complete evaluation of the participatory research approach applied in a research project. However, they are flexibly designed and can also be used interdependently (e.g. one instrument only). In addition, each element (e.g. each question of the qualitative interview guideline or the quantitative survey) of the Reflection Instruments is specifically coded in line with the change elements of the Impact Model. This means, questions can be selected in line with what changes the organizers of the participatory research activity are primarily intending to address. To make planning an evaluation of participatory research easier for research teams, on the interactive website each change element of the Impact Model can be selected and possibilities for its evaluation with different methodologies of the Reflection Instruments will be displayed.

9 See: https://ois.lbg.ac.at/en/projects/impact > "Tools" > "OIS Impact Reflection Instruments"

Figure 2 Timeline of Reflection Instruments



III.4 REFLECTION AND LEARNINGS

In this paragraph, we aim to provide our reflections on two aspects: first, we want to shed light on the advantages and challenges we faced in the process of applying the Theory of Change. Evaluators who aim to use this approach should gain insights into our experiences to better manage change processes in the future. Second, we want to highlight what the outcome of our process – the Impact Model and Reflection Instruments – can be used for in participatory research practice and where limitations apply.

III.4.1 REFLECTIONS ON MANAGING THE LBG IMPACT MODEL BASED ON A THEORY OF CHANGE PROCESS

Developing the Impact Model and Reflection Instruments using the Theory of Change approach had several advantages: first, the use of the Theory of Change structure enabled a co-creative development which represented diverse views. Perceptions on the (desired) changes and societal impact of participatory research can differ across stakeholder groups (researchers, for example, can wish for different societal changes than relevant for the involved public). The Theory of Change approach made it possible for all voices to be heard and considered. Enriching discussions and feedback loops fostered through the diverse representation of stakeholder groups, particularly refining wording and definitions, led to in-depth reflections and ultimately deeper understanding of the topic. Overall, this led to a more comprehensive model. Second, the Theory of Change approach allowed for a non-biased and bottom-up approach: Rather than top-down suggesting how participatory research leads to societal change and the subsequent risk of bias, we were able to develop expected pathways to impact from the beginning. Third, the Theory of Change approach, which has been tried and tested in different contexts, provided a lose structure for orientation, but still allowed for a highly individualized development of the contents: The input-output-outcome-impact structure enabled stakeholders with very different backgrounds (some had more, some less contact with the topic) to think about the pathways to impact in a structured way, but gave flexibility in determining what these changes may be.

Along the development process, we experienced some challenges: first, we needed to enable diverse stakeholder groups with different ties to the topic to speak the same language. We decided to split up the co-development process of the Impact Model into two workshops to allow us to keep a certain level of depth by grouping our stakeholders in (somewhat) similar groups (see chapter III.3.1). This stakeholder differentiation approach worked well: though the two workshops to develop the Impact Model were structured differently due to different target groups and their different experiences and touchpoints, their results – i.e. the change elements developed - complemented each other after some language and wording adaptations. A second challenge we experienced was inviting non-evaluation experts to contribute to developing evaluation methods in the third workshop. We solved it by asking participants for personal examples for each change level, so as to deduce ways and methods of evaluating these experienced changes without participants needing expertise in the field of evaluation. This strategy was successful and shows that involving non-experts of evaluation in developing evaluation methods can work with the right translation strategies. The added benefit of this approach was that we simultaneous uncovered several effects of the participatory research on participants in their practice example that had not yet been known to the research group Village.

In applying the Theory of Change to practice in the co-creative development of the Impact Model, we identified several success factors: first,

the mindset behind such a process is key. Unless organizers consciously decide to let go of power and control and transfer it to participants, they are likely to remain dominant of the agenda and will not benefit from the different perspectives involved. This requires a true understanding of the aims and the process, buy-in from all team members and challenges traditional leadership roles. Second, excellent facilitation skills are required. Our aim was complex and needed a careful translation to the context of our stakeholders, clear processes and structures and well-defined tasks so as to enable high-quality input of participants. We recommend using an external (expert) facilitator so as to enable the different stakeholders to speak at eye-level (in line with Plottu & Plottu 2011), while allowing the organizers of the process to step back. In our case, we highly benefitted from an external facilitator guiding the process with extensive knowledge on the Theory of Change, but from a different field and context. As the facilitator was not affiliated to any of the stakeholder groups present, they provided an outsider view, which prevented bias and conflict. Third, a proper Theory of Change process requires resources and dedication. In our case we had to pay meticulous attention to detail in the use and connotation of certain words when naming and defining each impact element. This reflection process required us to regularly step back, review and revisit the Impact Model at a later stage to ensure a balanced view. This took more time than anticipated. These implementation costs should be considered in advance. Finally, we used the Theory of Change approach to develop an Impact Model and Reflection Instruments that would allow a comparison between different participatory research projects on the basis of the participatory research approaches applied. Applying the framework to a specific research project, however, would mean that a direct comparison of evaluation results across different research projects is made more difficult, meaning comparability and transferability can be impaired. This tailored approach makes sense, as the Theory of Change approach particularly enables an impact orientation of a research project, which can differ strongly even within a specific field. Yet, being aware of this downside should allow a proper consideration if the purpose of the Theory of Change approach fits the aim. To enable an impact-oriented view, a project could also apply certain principles of a Theory of Change structure, rather than fully embracing the whole process.

III.4.2 PRACTICAL USE

Researchers can use the Impact Model to plan the societal impact of their participatory research activities. The model allows to explore different change levels in participants, which ultimately lead to research uptake and societal impact of participatory research. The Reflection Instruments, at the same time, provide a set of different methods and tools to evaluate these societal changes. They can be used as a complete set or selectively (e.g. by merely integrating specific question items into existing evaluation instruments). Moreover, the planning tool – the interactive Impact Model – allows to explore how each of these changes could be evaluated individually. Therefore, researchers can select the anticipated changes most important to them. Each change is backed with various evaluation suggestions using different methodologies. Researchers can thus easily integrate individual evaluation items into existing evaluation instruments and structures (e.g. mandatory evaluations), making use of them for their own purpose. As they are deduced from real-life practice examples, they are easy to implement into practice. Finally, the Impact Model and Impact Reflection Instruments offer a new angle on theory-based evaluation: rather than focused on a specific research project or a research discipline, the Impact Model and Impact Reflection Instruments systematically uncover societal changes due to the participatory research approach applied. They therefore provide increased comparability to participatory research across different disciplines and fields.

The Impact Model and Reflection Instruments are designed for researchers and research teams as a useful tool to plan and establish feedback mechanisms in a simple and practical way. Reflection Instruments do not replace a thorough and academically rigid external evaluation, but should rather provide immediate feedback through monitoring throughout and after the participatory research activities. Researchers should be enabled to strategically plan for the societal impact of their participatory research activities, as well as to autonomously receive feedback. This should empower them to continuously strive for and improve the societal impact orientation of their participatory research so as to increase the chances of research uptake and societal impact.

III.4.3 LIMITATION AND NEXT STEPS

The strong orientation towards practicability and usability of the Impact Model and the Reflection Instruments explains why their development relied largely on a bottom-up approach based on practical experience and examples. While the Theory of Change approach provided a strong structure, we purposefully did not rely on existing measurement frameworks in literature, but deliberately decided on a purely bottom-up approach to develop Reflection Instruments and its indicators. Although the literature is inconsistent, this approach could create tension with existing frameworks. What is more, while the input of stakeholders involved in the development of the Impact Model largely overlapped, the quantity of stakeholders involved was limited. Finally, the practice orientation towards a specific project in the development of the Impact Model and Reflection Instruments facilitated the process, but this close practice-oriented view may have led to blind spots in fields not represented in the co-development process. The Impact Reflection Instruments were initially set out to be applied and tested in a pilot run in cooperation with the LBG mental health research group Village. However, due to the developments of the Covid-19-pandemic, these plans had to be put on hold. Thus, it remains to be seen whether the Reflection Instruments can be useful and applicable to different participatory research projects in different contexts.

To validate the applicability of the Impact Model and Reflection Instruments in other contexts, we therefore aim to test them in different participatory research projects across various fields and disciplines, contexts and operating under different scopes and time spans. The aim is to identify whether the Reflection Instruments provide valuable feedback for a variety of application contexts and to refine them further to respond to varying demands that may arise in the application process.

IV CONCLUSION

Societal impact of research receives increasing attention across different national research landscapes as the wider benefits and impacts of research activities for societies at large come into scope. Despite disciplinary differences in impact definitions and evaluations, there is a general

and increased understanding across basic, theoretical and more practically, applied research fields that societal relevance of research activities should be strived for without compromising research autonomy. In fact, finding a balance between the autonomy of science and its aspiration to be societally relevant lies at the heart of most impact debates. Also, collaborative and participatory research work and involvement activities are increasingly considered central for leveraging societal research impact and the significance of building and facilitating sustainable relationships between researchers and other non-academic stakeholders is being more and more acknowledged.

To make societal impact reflections and evaluation processes an integral part of participatory research processes, different incentives have to be set that consider the time, costs and expertise needed. Instead of rewarding individual researchers and projects for achieving impact based on narrow indicators, funding schemes should be directed towards creating research environments and organizational structures that foster participatory endeavors, relationships building and regular exchange with different stakeholder groups. On a more individual level, researchers should receive training and methodological tools and expertise in order to build capabilities for participatory research processes, including impact evaluation and reflection.

The Impact Model and Reflection Instruments were designed and co-created as easy to use and adaptable tools to support researchers navigate through their participatory research projects, to reflect and anticipate some of the impacts arising from them and to check and evaluate the feedback from both co-researchers and other stakeholders involved in the project. While the focus here is on participatory research projects, we also strongly encourage researchers in different research contexts, projects and across disciplinary fields to use the Impact Model and Reflections Instruments in order to see how it can be utilized and adapted differently and also to derive more generalizable information and recommendations for further enhancements.

While there are different evaluation approaches to uncover the societal impact of participatory research, no standardized framework has been defined as of yet. The Theory of Change approach, however, is commonly used for societal impact evaluation and participatory research. We introduced a co-creative and bottom-up development of an Impact Model based on the Theory of Change approach to complement existing literature from a practice- and stakeholder-based perspective. In addition, we shed light on the application of a Theory of Change approach in practice. We shared advantages of the process, challenges we faced and how we addressed them, as well as essential factors for successful use of the Theory of Change approach. These reflections should help evaluators who aim to make use of the Theory of Change in contexts relevant to their work guide their application process.

REFERENCES

Arnstein, S. R. (1969). A ladder of citizen participation. Journal of the American Institute of Planners, 35(4), 216-224.

Barber R, Boote JD, Parry GD, Cooper CL, Yeeles P, Cook S. (2012). Can the impact of public involvement on research be evaluated? A mixed methods study. Health Expectations, 15.

Baum, F., MacDougall, C., Smith, D. (2006). Participatory Action Research. Journal of Epidemiology and Community Health, 60, pp.854-857. Beck, S., Bergenholtz, C., Bogers, M., Brasseur, T.M., Conradsen, M.L., Di Marco, D., Distel, A., Dobusch, L., Dörler, D., Effert, A., et al. (2020) The open innovation in science research field: A collaborative conceptualisation approach. Industry and Innovation.

Belcher, B.M., Davel, R., Claus, R. (2020). A refined method for theory-based evaluation of the societal impacts of research. MethodsX, 7.

Boivin, A., L'Espérance, A., Gauvin, F.P., Dumez, V., Macaulay, A.C., Lehoux, P., et al. (2018). Patient and public engagement in research and health system decision making: A systematic review of evaluation tools. Health Expectations, 21(6), pp.1075–84.

Bonney, R., Cooper, C.B., Dickinson, J., Kelling, S., Phillips, T.B., Rosenberg, K.V., Shirk, J. (2009). Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. Bioscience, 59(11), p.977–984.

Bornmann, L. (2012). What is societal impact of research and how can it be assessed? A literature survey. Journal of the American Society for Information Science and Technology, 64(2), pp. 217-233.

Bührer, S., Yorulmaz, M., Feudo, F., Reichmann, S., Gomez, L. B., Kinnula, M., van Laar, M., Balestrini, M., Stack, S., Rüfenacht, S., Riemenschneider, D. (2021). Measuring the Impact of RRI. Lessons learned from Pilot Action 1 of the Social Lab No. 15 SwafS. NewHorrizon

Caracelli, V.J., Greene, J.C. (1997). Crafting Mixed-Method Evaluation Designs. New Directions for Evaluation, 74, pp.19-32

Cargo, M., Mercer, S. L. (2008). The value and challenges of participatory research: Strengthening its practice. Annual Review of Public Health, 29, pp- 325-350.

Cousin, J.B., Whitmore, E. (1998). Framing participatory evaluation. In: Understanding and practicing participatory evaluation. (Ed) Whitemore, E., New directions for evaluation (Vol. 80, pp. 5-23). San Francisco: Jossey-Bass

Donovan, C., Hanney, S. (2011). The 'Payback Framework' explained. Research Evaluation, 20(3), pp.181–183.

Felt, U., Fochler, M. (2018). The societal impact of social science knowledge in Austria: Impact pathways, measurements, potential. An explorative study, commissioned by the Austrian Council for Research and Technology Development. Retrieved from https://www.rat-fte.at/files/

rat-fte-pdf-en/documents/publications/190124_Societal_Impact_Felt_Fochler%202018.pdf

Gallivan J., Kovacs Burns K., Bellows M., Eigenseher C. (2012). The many faces of patient engagement. Journal of Participatory Medicine 32(4). Retrieved from http://www.jopm.org/evidence/research/2012/12/26/the-many-faces-of-patient-engagement/

Greenhalgh, T., Raftery, J., Hanney, S., Glover, M. (2016). Research impact: a narrative review. BMC Medicine, 14(78).

Hayes, H., Buckland, S., Tarpey, M. (2012). Briefing notes for researchers: public involvement in NHS, public health and social care research. Eastleigh, England: INVOLVE.

Hoekstra, F., Mrklas, K. J., Khan, M., McKay, R. C., Vis-Dunbar, M., Sibley, K. M., Nguyen, T., Graham, I. D., SCI Guiding Principles Consensus Panel, Gainforth, H. L.(2020). A Review of reviews on principles, strategies, outcomes and impacts of research partnerships approaches: A first step in synthesizing the research partnership literature. Health Research Policy and Systems, 18(1), pp.1-23.

Kieslinger, B., Schäfer, T., Heigl, F., Dörler, D., Richter, A., Bonn, A. (2018). Evaluating citizen science: Towards an open framework. In Citizen Science: Innovation in Open Science, Society and Policy. (Ed) Hecker, S., Haklay, M., Bowser, A., Makuch, Z., Vogel, J., Bonn, A. UCL Press: London, UK, 2018; pp.81–95.

Klein, J.T. (2013). The Transdisciplinary Moment(um). Integral Review 9, 189–199.

Muhonen, R., Benneworth, P., Olmos-Penuela, J. (2019). From productive interactions to impact pathways: Understanding the key dimensions in developing SSH research and societal impact. Research Evaluation, 29(1).

Oancea, A. (2013). Interpretations of Research Impact in Seven Disciplines. European Educational Research, 12(2).

Owen, R., Pansera, M. (2019). Responsible Innovation and Responsible Research and Innovation. In: Handbook on Science and Public Policy. (Ed) Simon, D., Kuhlmann, S., Stamm, J., Canzler, W. Edward Elgar publishing.

Patton, M. Q. (1994). Developmental Evaluation. Evaluation Practice, 15(3), pp.311–319.

Pawson, R. A. Y., Tilley, N. (1994). What Works in Evaluation Research? The British Journal of Criminology 34(3), pp.291-306.

Phillips, T., Porticella, N., Constas, M., Bonney, R. (2018). A Framework for Articulating and Measuring Individual Learning Outcomes from Participation in Citizen Science. Citizen Science: Theory and Practice, 3(2), p.3.

Phipps, D., Cummings, J., Pebler, D., Craig, W., Cardinal, S., (2016). The Co-produced Pathway to Impact Describes Knowledge Mobilization Processes. Journal of Community Engagement and Scholarship, 9(1).

Plottu B., Plottu E. (2011). Participatory Evaluation: The Virtues for Public Governance, the Constraints in Implementation. Group Decision and Negotiation, 20 (6), pp.805-824.

Reed, M.S. (2016). The Research Impact Handbook, Fast Track Impact

Reed, M.S., Duncan, S., Manners, P., Pound, D., Armitage, L., Frewer, L., Thorley, C., Frost, B. (2018). A common standard for the evaluation of public engagement with research. Research for All, 2(1), pp 143–162.

Reed, M.S., Ferré, M., Martin-Ortega, J., Blanche, R., Lawford-Rolfe, R., Dallimer, M., Holden, J. (2021). Evaluating impact from research: A methodological framework. Research Policy, 50(4).

Rymer, L. (2011). Measuring the impact of research—The context for metric development. Turner, Australia: The Group of Eight.

Smith, K., Bandola-Gill, J., Meer, N., Stewart, E., Watermeyer, R. (2020). The Impact Agenda. Controversies, Consequences and Challenges. Policy Press.

Smits, P.A., Champagne, F. (2008). An Assessment of the Theoretical Underpinnings of Practical Participatory Evaluation. American Journal of Evaluation, 29(4), pp.427-442.

Spaapen, J., van Drooge, L., Propp, T., Shinn, T., Marcovich, A. (2011). Social Impact Assessment Methods for research and funding instruments through the Study of Productive Interactions between science and society. SIAMPI final report, http://www.siampi.eu/Content/SIAMPI_Final%20report.pdf, www.siampi.eu

Staley K. (2015). Is it worth doing? Measuring the impact of patient and public involvement in research. Research Involvement and Engagement. Turnbull, B. (1999). The mediating effect of participation efficacy on evaluation use. Evaluation and Program Planning, 22, pp.131-140.

UKRI website (2021). Defining Impact. https://www.ukri.org/councils/esrc/impact-toolkit-for-economic-and-social-sciences/defining-impact/ Accessed on 27 October 2021.

Van den Akker, W., Spaapen, J. (2017). Productive interactions: societal impact of academic research in the knowledge society, LERU position paper. https://www.leru.org/publications/productive-interactions-societal-impact-of-academic-research-in-the-knowledge-society

Vaughn, L. M., & Jacquez, F. (2020). Participatory Research Methods – Choice Points in the Research Process. Journal of Participatory Research Methods, 1(1).

Wehn, U., Gharesifard, M., Ceccaroni, L., Joyce, H., Ajates, R., Woods, S.M., Bilbao, A., Parkinson, S., Gold, M., Wheatland, J. (2021). Impact Assessment of Citizen Science: state of the art and guiding principles for a consolidated approach. Sustainability science, 16, pp.1683-1699.

AUTHORS

MAGDALENA WAILZER
Open Innovation in Science Center
Ludwig Boltzmann Gesellschaft
Nussdorferstraße 64, 1090 Wien
E: magdalena.wailzer@lbg.ac.at

LAURA SOYER
Open Innovation in Science Center
Ludwig Boltzmann Gesellschaft
Nussdorferstraße 64, 1090 Wien
E: laura.soyer@lbg.ac.at

KEYWORDS: Societal impact, participatory research, impact evaluation, theory of change

THE IMPACT OF RESPONSIBLE RESEARCH AND INNOVATION (RRI) - A CO-CREATED TEMPLATE WITH A COMPILATION OF THE SCIENTIFIC, SOCIETAL AND ECONOMIC IMPACTS OF RRI

MERVE YORULMAZ AND SUSANNE BÜHRER

DOI: 10.22163/fteval.2022.548

ABSTRACT

In this paper, we show how the policy concept Responsible Research and Innovation (RRI) intends to offer added value to scientific research processes and its manifold results and how these anticipated benefits can be measured. We set out to address the recognized challenge of impact monitoring for projects working with RRI by developing a template that can pay attention to the so far hardly considered outputs, outcomes and impacts of a particular RRI project. Its usage goes beyond monitoring and evaluating purposes at the end of a project, as our proposed set of indicators can provide scholars and practitioners with guidance and inspiration in the early design or implementation phase of a project. From a policy-making point of view, this paper also highlights that developing monitoring and evaluation systems can significantly profit from stakeholder engagement and co-creation approaches, adding a bottom-up perspective to top-down suggestions of the research funders.

EMBEDDING RESPONSIBILITY IN EUROPEAN RESEARCH AND INNOVATION PROJECTS

In recent years, various societal drivers have increased the pressure on the science system to legitimize the use of public funds. Typically, it is no longer deemed sufficient to achieve goals intrinsic to research, such as contributing to the development of theory and methods or achieving knowledge gains. Instead, the contribution that research makes to solving problems matters, especially for major societal challenges. This debate is strongly linked to keywords such as Sustainable Development Goals, Societal Impacts, a "New social contract of science", and also in particular concepts that aim to improve the relationship between science and society (Gibbons, 1999; Demeritt, 2000; Reale et al., 2015; Martin, 2011; von Schomberg, 2013).

As regards the latter, the concept of responsible research and innovation (RRI) that was initiated and promoted by the European Commission (EC), has gained particular attention. The strengthened and improved relationship between science and society is at the core of RRI which aims to embed responsibility as a core value in research and innovation processes and cultures (European Commission, 2014; Lindner et al., 2016; Bogner et al., 2015). Generally speaking, RRI intends to create a new and improved relationship between science and society. The EC defines RRI as "a process where all societal actors (researchers, citizens, policy makers, business) work together during the whole Research and Innovation (R&I) process in order to align R&I outcomes with the values, needs and expectations of European society" (von Schomberg 2013). RRI has also been defined as "... a transparent, interactive process in which societal actors and innovators become mutually responsive to each other with a view on the ethical acceptability, sustainability and societal desirability of the innovation process and its marketable products." (Call FP7-SiS-2012-1). Scholars like Stilgoe et al. (2013) emphasize four integrated dimensions that characterize RRI, namely anticipation, reflexivity, inclusion, and responsiveness.

The Commission has promoted the concept of RRI in a number of ways: First, through research funding under the various EU Framework Programmes. Here, the naming of the relevant programme pillars underscores the changing understanding of science-society relations that have evolved: from Science and Society (SaS, Framework Programme 6) to Science in Society (SiS, Framework Programme 7) all the way up to Science with and for Society (SwafS, Horizon 2020). While "SaS" could still be interpreted as two separate subsystems interacting with each other, "SiS" already implies close interpenetration. Finally, "SwafS" emphasizes both the normative (for) and the participatory (with) components. Horizon Europe integrates the science-society relationships, namely to engage and involve citizens and civil society organisations, in different regards, as part of the Horizon Europe Regulation and Specific Programme, as an excellence criterion and as element in the key impact pathways.

In the following, we show that normative orientation is associated with specific challenges, as it involves a cognitive bias towards benefits, without at the same time pointing out potential negative effects.

1 The Higher Education Institutions and Responsible Research and innovation (HEIRRI) project aimed to bring the concept of RRI into the educational system (https://cordis.europa.eu/project/id/666004).

ESTABLISHING AN RRI ECOYSTEM

With the help of research funding, an RRI community of experts has emerged with numerous projects that, in addition to analytical issues such as the need for structural change in the science system (HEIRRI[1], JERRI[2], RRI Practice)[3] and altered governance structures (Resagora[4]), has developed very practical approaches to bringing RRI into practice (RRI Tools[5], FOTRRIS[6], to mention just a few). On the other hand, the EU has specifically set up expert panels and commissioned services to advance the topic of RRI.

Two initiatives deserve special attention: the expert group on "Indicators for promoting and monitoring Responsible Research and Innovation"[7] and the MoRRI project[8]. The MoRRI project aimed to monitor the evolution and benefits of RRI. In doing so, the study team developed numerous indicators for the so-called RRI keys Public Engagement, Ethics and Governance, Gender Equality, Science Education and Open Access (Peter et al., 2018). The starting point of MoRRI was an intervention logic. Accordingly, a distinction was made between inputs ("responsible practices"), outputs resulting directly from them, and longer-term outcomes or impacts. The indicators referred exclusively to the national level, even though they were often created from aggregated data at the individual or organizational level. In addition, so-called "benefit indicators" were also developed within the framework of MoRRI, which referred to the following dimensions: scientific, economic, democratic, and societal.

The EC then required in the Horizon 2020 work programmes (WP) that project applicants apply the MoRRI indicators: "Several WP18-20 topics specify indicators which applicants should work towards, notably from the Sustainable Development Goals and from the study Monitoring the Evolution and Benefits of Responsible Research and Innovation (MoRRI)" (European Commission, Horizon 2020 Work Programme 2018-2020, p. 8). However, due to the specific objective, approach and results of MoRRI – developing indicators that refer to the country level – it was hard to apply them at a project level.

Thus, two developments came together that posed significant challenges on the Horizon 2020 project applicants and beyond: How can researchers address the increasing demand of public research funders to demonstrate the impact of their research? How can RRI-related impacts be measured?

MEASURING THE IMPACTS OF RRI - A CHALLENGING ENDEAVOR

In recent years, there have been several attempts to measure the impacts of RRI in the context of EU projects and programmes. The EU project RI-Paths, for example, has compiled various indicators to measure societal impacts of research infrastructures. These indicators show various references to the RRI concept as developed by the EC, focusing on different so-called RRI keys (Public Engagement, Ethics and Governance, Gender Equality, Science Education and Open Access). Specifically, the RI pathway indicators mention the aspects of open access, public engagement, science communication and gender equality (Helman et al., 2020, p. 16). Moreover, the EU Expert Group on Monitoring the EU Framework Programmes (2018) has proposed "citizen engagement" as part of the Key Impact Pathways. The reason why citizen engagement is understood as a relevant impact dimension is that participatory processes are considered as important for legitimacy, accountability and transparency of research and innovation (ibid.).

However, in contrast to these practical suggestions and recommendations, there are still critical voices that warn against specifying indicators too quickly, for example the fteval Working Group on Impact Assessment: "a limited set of indicators will not do justice to the multidimensional character of the undertaking. The relationship between science and society must be considered in all of its dimensions, bearing in mind that this relationship is embedded in a complex system of formal and informal interactions that are open to change over time. RTI policy interventions to create and maintain such interactions are themselves important mechanisms for opening a space to shape the relationship between science and society and define collaboratively the benefits of this relationship"[9]. Nonetheless, despite these efforts, there is still need for a framework to impact monitoring for projects working with RRI that is more comprehensive and sensitive to the specificities of RRI impacts.

THE MONITORING / EVALUATION TEMPLATE

To address this challenge and need, we propose a framework of measuring the short-, medium- and long-term results of RRI that revolves around the question how the scientific, economic, societal and demo-

2 The Joining Efforts for RRI project (JERRI) aimed at fostering RRI transition in Europe by developing and testing good RRI practices in pilot cases, for a further upscaling among the RTOs in the EU28 (https://cordis.europa.eu/project/id/709747)

3 RRI-Practice project brought together a unique group of international experts in RRI to understand the barriers and drivers to the successful implementation of RRI both in European and global contexts (https://cordis.europa.eu/project/id/709637)

4 The Resagora project (Responsible Research and Innovation in a Distributed Anticipatory Governance Frame. A Constructive Socio-normative Approach) developed a normative and comprehensive governance framework for Responsible Research and Innovation (RRI) (https://cordis.europa.eu/project/id/321427).

5 The RRI Tools project developed a training and dissemination toolkit on RRI (https://cordis.europa.eu/project/id/612393).

6 Fostering a Transition towards Responsible Research and Innovation Systems (FOTRRIS) developed new governance practices to foster RRI policies and methods in research and innovation systems (https://cordis.europa.eu/project/id/665906).

7 The expert group was chaired by Roger Strand, professor at the Centre for the Study of the Sciences and the Humanities and the CCBIO, University of Bergen.

8 The final reports can be found here: https://super-morri.eu/morri-2014-2018/

9 Blog post, page 9, https://www.fteval.at/content/home/news/ag_impact_results/AG-Impact_G2-Sci-Soc_Blogpost.pdf

cratic outputs, outcomes and impacts of RRI can be defined and empirically collected at the project level. The work presented in this paper can be understood as a bottom-up based response to the top-down decision of the EC to use the MoRRI indicators as an element for project applications.

CO-CREATION IN INNOVATIVE SPACES FOR SOCIAL EXPERIMENTATION

The monitoring and evaluation template, which is presented in further detail in the following sections, is the result of a two-year bottom-up co-creation process between international R&I stakeholders from academia, education and business. It took place within a novel social experimentation format in the context of the Horizon 2020 project NewHoRRIzon ("Excellence in science and innovation for Europe by adopting the concept of Responsible Research and Innovation", 2017-2021)[10]. This project aimed to promote the integration of RRI into European, national and local R&I practice and EU funding. Methodologically, it was built around 19 social labs – "platforms for addressing complex societal challenges" (Hassan, 2014, p. 3) – each of which is dedicated to a different section of Horizon 2020, the past European Framework Programme for Research and Innovation.

The template originated from a so-called *pilot action*[11] under the umbrella of the social lab dedicated to the SwafS programme line. According to the understanding developed in the NewHoRRIzon context, such pilot actions can be understood as activities that are jointly implemented by a group of stakeholders in form of social experiments that aim to tackle a specific societal challenge by systematically integrating aspects of RRI. They emerge fully bottom-up by the initiative of stakeholders interested in or affected by the identified challenges and aim at practical implementation, thus perfectly responding to the recognized deficiency and impracticability of the original set of MoRRI indicators. The specific value of this activity is thus, inter alia, that it is tailored to the aforementioned need for Horizon 2020 projects to develop their impact sections along the MoRRI indicator framework.

PROCESS, STRUCTURE AND PURPOSE OF THE TEMPLATE

From a practical point of view, the template emerged from three (physical, pre-Corona) working meetings between 12 committed R&I stakeholders who built a pilot action in order to discuss the template's overarching structure, suitable categorizations and formulation of individual indicators in an ongoing process of reflection and refinement of its objectives, practicability and overall utility. This work was done against the background of the pilot group's individual needs, experiences, and expertise. The comprehensive consideration of the interests and needs of the pilot action members was ensured by appropriate moderation and the noteworthy group diversity in terms of disciplinary background, country of origin, age, gender, and expertise in one of the RRI keys. In fact, great stakeholder diversity is considered a key pillar of the social lab methodology and a means to ensure divergence in perspectives to stimulate creative and critical thinking (Blok, 2019) in light of the defined task's complexity and significance. In this sense, the debate in the pilot work was strongly shaped by questions around the long-term materialization of impacts (Wittmann et al., 2021), or the non-linear, context-sensitive and emergent nature of impact pathways, to name only a few methodological, conceptual or practical challenges of impact assessment.

This process resulted in the development of a multi-page template that is divided into five major sections: The first two sections cover a set of questions and criteria that are related to the project's nature and the consideration of RRI dimensions in the project's design, tailored to each of the five RRI dimensions. Figure 1 presents an exemplary overview of the statements that ask for the systematic consideration or integration of aspects tailored to the RRI key "Public Engagement / Citizen Science".

The last three sections of the template (see Figure 2) capture the three main dimensions for impact fields, i.e., 1) scientific, 2) economic and 3) societal and/or democratic impacts of RRI. Each of these are presented in a matrix that lists the respective indicators (vertically) along a 4-point scale (horizontally). In alignment with the template's fundamental aim to better account for temporality and the multidimensionality of the effects of RRI, the indicators are subdivided into short-, medium- and longer-term impacts. The final indicator list is comprised of a modified set of existing MoRRI indicators and to a large extent of new indicators developed by the pilot group (for more details see Bührer et al., 2021). Figure 2 provides an overview of the indicators for the democratic / societal effects of RRI.

As regards the core purpose of the template, it aims to show and communicate the benefits and impacts of RRI in a systematic manner to academic and non-academic audiences by equipping them with a practical, multi-purpose template that can inspire and navigate them throughout the entire lifecycle of an RRI project. More concretely, potential users of and target groups for the template are researchers and practitioners interested in measuring their project impacts with the help of indicators that are inspired by the MoRRI indicators, but which go beyond and are jointly developed by a group of interested stakeholders, and thus more user-friendly. Its usage goes beyond monitoring and evaluating purposes at the end of a project since the lists of indicators and descriptors can provide project managers with guidance and inspiration in the early design or implementation phase of a project.[12]

10 https://newhorrizon.eu/social-labs/

11 Requirements for pilot actions in the context of Social Labs of the NewHoRRIzon project were that their objectives can be realistically achieved in terms of available time, money and capacities, that they are linked to the interests of the stakeholders engaged in the pilot action and that it has the potential to be of interest for other stakeholders beyond the own pilot action as well.

12 More information on the Pilot Action and the template can be found here: https://newhorrizon.eu/social-lab-15-pilot-action-1/

Figure 1: The role of RRI in the project with the RRI key example Public Engagement

II. The role of RRI

Are the following aspects systematically taken into account in your project?
(Answer categories: Yes / No / don't know)

		Yes	No	I don't know
Public Engagement / Citizen Science	☐ I inform non-academics about my results through e.g. public lectures, writing popular science books, publishing articles in newspapers / magazines, blogs	☐	☐	☐
	☐ I involve citizens in the following phase(s) of my research by:	☐	☐	☐
	☐ definition of content and aims	☐	☐	☐
	☐ conducting the research (data collection, data analysis)	☐	☐	☐
	☐ discussing the consequences of research and / or its application	☐	☐	☐
	☐ Communicating and disseminating the results of the project	☐	☐	☐
	☐ Commercialisation / Exploitation of results	☐	☐	☐
	☐ I actively consider how my research and innovation results will be perceived and used	☐	☐	☐
	☐ I work with people who specialise in dialogue with citizens and civil society (e.g. professional mediator, communication company, science museums)	☐	☐	☐
Does your research and innovation process foresee a systematic inclusion of stakeholder groups outside academia?		☐ Yes, an active involvement of previously marginalised or disenfranchised actors is foreseen ☐ If so, which groups are involved? (please specify) _____		
		☐ Yes, the introduction of previously excluded perspectives and knowledge sources into R&I is foreseen: ☐ If so, what are the concrete instruments to do so? (please specify) _____		

Figure 2: Template for democratic / societal benefits of RRI

V. Societal and democratic impacts / benefits of RRI

Do / Did you expect or observe any of the impacts benefits listed below when practicing RRI?

		I expect a respective impact / benefit	I do not expect such kind of an impact / benefit	I have already observed such an impact / benefit	I don't know / not applicable
Short -term outputs	Increased researchers' awareness of potential negative effects on citizens (precautionary principle)	☐	☐	☐	☐
	Broaden problem framing	☐	☐	☐	☐
	Increase science capital by increasing skills and knowledge among citizens and communities (regardless of your legal status)	☐	☐	☐	☐
	Evidence on the positive effects of science education	☐	☐	☐	☐
	Increased awareness of unconscious / personal biases	☐	☐	☐	☐
	Outreach to disadvantaged groups	☐	☐	☐	☐
Midterm outcomes	Increased researchers' awareness of potential negative effects on citizens (precautionary principle)	☐	☐	☐	☐
	Broaden problem framing	☐	☐	☐	☐
	Increase science capital by increasing skills and knowledge among citizens and communities (regardless of your legal status)	☐	☐	☐	☐
	Evidence on the positive effects of science education	☐	☐	☐	☐
	Increased awareness of unconscious / personal biases	☐	☐	☐	☐
Long-term impacts	Enhancement of Knowledge through access to knowledge	☐	☐	☐	☐
	Behavioural change among citizens	☐	☐	☐	☐
	Improved scientific citizenship and trust in science	☐	☐	☐	☐
	Improved education system	☐	☐	☐	☐
	More inclusive societies	☐	☐	☐	☐
	More equitable societies	☐	☐	☐	☐

CONCLUSION

The previous sections have shown that within the framework of a social lab, it is possible to further develop existing evaluation and monitoring systems in a way such that they reflect the needs of the users well or at least better. The starting point for our work was the finding that the requirements for applicants to Horizon 2020 projects to use MoRRI indicators was impractical for a number of reasons. Together with project representatives[13] from a wide range of fields and with a variety of scientific and institutional backgrounds, we succeeded in further developing the existing indicators and increasing their practicability. In doing so, we contributed to the recognized challenge of impact monitoring for projects working with RRI by developing a theoretically sound and practical, multi-purpose template that can pay attention to the so far hardly considered outputs, outcomes and impacts of a certain RRI project.

We are aware that our attempts to develop an easy-to-use template to capture and measure impacts is not without limitations. We recognize that this may lead to an oversimplification of processes that are much more complex, non-linear and dynamic in nature. For example, it should be noted that we still followed an ideal intervention logic. This means that we differentiate between short-, medium- and long-term results, even if it is well known that such kind of linear models do not sufficiently reflect the complexity and feedback loops of R&I activities. However, this is still the common way to operate in R&I program monitoring and evaluation activities, as can be seen in the EU Key Impact Pathways-approach as well. Another challenge that we have certainly not yet fully addressed is the vagueness of the concept of societal impacts. Here, several authors have demonstrated that there is still no consensus on what exactly is meant by this impact type (Reale et al., 2015; Bornmann, 2013; Smit and Hessels, 2021; Williams, 2020; Sivertsen and Meijer, 2020; Muhonen et al., 2020).

Furthermore, it needs to be stated that our indicators and descriptors only depict positive effects. Although the proposed indicators can always be interpreted in two directions – as the presence or absence of a benefit – the entire concept of the template is designed to show a positive contribution of science to society. This follows a policy agenda that, if one recognizes the importance of improved science and society relations in general and specific RRI activities in particular, uses such data collection specifically to demonstrate the contribution of science to generate public value and to advance RRI. This is, therefore, a normative and less "objective" approach to measuring impact.

Finally, one needs to be wary of the danger that such a template and resulting indicators can also foreclose learning and lead to unwanted bureaucracy.

However, for the ease of use and to ensure connectivity with other monitoring and evaluation systems (in particular, the Key Impact Pathways used for Horizon Europe), we believe that this approach is justified and provides new avenues for future work and the further refinement of the template. Finally, although we learned that it is rather difficult and complex to work on this topic, it is worthwhile to do it because many researchers in EU funded projects need to make at least some use of indicators to show the benefits and impacts of RRI.

A key lesson that we can draw from the work on developing an evaluation system for RRI in the context of a social lab is, in short, that despite all the effort and weaknesses, the approach is worthwhile. The participatory approach complements an expert-based top-down exercise typically used by the EC, but also by many national funders, in developing evaluation systems. In concrete terms, this means that the research funder invites a selection of high-level experts to develop a scientifically sophisticated evaluation system, but its implementation is then top-down, without adequately reflecting the realities and needs of the users of this system. As we know from current discussions about the requirements for evaluating transformative policy approaches, and as such we can in principle also understand RRI, a more intensive stakeholder involvement in all the evaluation phases including the design phase is needed (Molas-Gallart et al., 2020). The above presented results show what this could look like in practice, acknowledging that only the (generous) project funding from Horizon 2020 allowed us to start such an intense collaboration and co-creation process.

REFERENCES

Blok, V. (2019): From participation to interruption: Toward an ethics of stakeholder engagement, Participation and Partnership in Corporate Social Responsibility and Responsible Innovation. (https://www.researchgate.net/publication/330171626_From_participation_to_interruption_Toward_an_ethics_of_stakeholder_engagement_Participation_and_Partnership_in_Corporate_Social_Responsibility_and_Responsible_Innovation/link/5f05c4c7a6fdcc4ca455e788/download).

Bührer, S.; Yorulmaz, M.; Feudo, F.; Reichmann, S.; Barbosa Gomez, L.; Kinnula, M., van Laar, M.; Balestrini, M.; Stack, S.; Rüfenacht; S. and D. Riemenschneider (2021): Measuring the Impact of RRI. Lessons learned from Pilot Action 1 of the Social Lab No. 15 SwafS. (https://newhorrizon.eu/wp-content/uploads/2021/05/SL15-Pilot-Action-1-Measuring-the-Impacts-of-RRI-Results_17-5-21.pdf).

Bogner, A.; Decker, M.; Sotoudeh, M. (eds.) (2015): Responsible Innovation. Neue Impulse für die Technikfolgenabschätzung? (Gesellschaft Technik Umwelt, Band 18). Nomos-Verlag. Baden-Baden ISBN: 9783848728824.

Bornmann, L. (2013): What Is Societal Impact of Research and How Can It Be Assessed? A Literature Survey. Journal of the American Society for Information Science and Technology, 64(2):217–233, 2013 (https://doi.org/10.1002/asi.22803) .

Demeritt, D. (2000): The New Social Contract for Science: Accountability, Relevance, and Value in US and UK Science and Research Polic,. Antipode 32:3, 2000, pp. 308–329 (https://onlinelibrary.wiley.com/doi/epdf/10.1111/1467-8330.00137).

13 The members of the pilot actions were mobilised by an EU-wide call for participation launched by EU project officers. The communication took place via the project coordinators, some of whom recommended members of their consortium for participation in the pilot action.

European Commission, Directorate-General for Research and Innovation (2014): Responsible Research and Innovation – Europe's ability to respond to societal challenges. Brussels (https://op.europa.eu/en/publication-detail/-/publication/2be36f74-b490-409e-bb60-12fd438100fe/language-en).

European Commission, Directorate-General for Research and Innovation (2018): Flecha, R., Radauer, A., Besselaar, P., Monitoring the impact of EU Framework Programmes : expert report, Publications Office (https://data.europa.eu/doi/10.2777/518781; European Commission, Directorate-General for Research and Innovation, Flecha, R., Radauer, A., Besselaar, P., Monitoring the impact of EU Framework Programmes : expert report, Publications Office, 2018, https://data.europa.eu/doi/10.2777/518781).

European Commission, Horizon 2020 Work Programme 2018-2020, 16. Science with and for Society (https://ec.europa.eu/research/participants/data/ref/h2020/wp/2018-2020/main/h2020-wp1820-swfs_en.pdf).

fteval Working Group on Impact Assessment. Impact of RTI-Policy on the relationship between science and society. Blog Post (v 1.22, 8/2/2021) (https://www.fteval.at/content/home/news/ag_impact_results/AG-Impact_G2-Sci-Soc_Blogpost.pdf).

Gibbons, M. (1999): Science's new social contract with society. Nature, Vol 22, C81-C84 (https://www.nature.com/articles/35011576.pdf).

Hassan, Z. (2014): The Social labs revolution: A new approach to solving our most complex challenges. Berrett-Koehler Publishers (https://ssir.org/books/excerpts/entry/the_social_labs_revolution_a_new_approach_to_solving_our_most_complex_chall).

Helman, A.; Barberis, Matias; Vignetti, Silvia; Jessica Catalano; Griniece, Elina; Kroll, Henning; Zenker, Andrea; Martin, Corinne (2020): Deliverable 5.1. Title: Validated IA Model. Research Infrastructure Impact Assessment Pathways. (https://ri-paths.eu/wp-content/uploads/2018/03/D5.1_Validated-IA-model.pdf).

Lindner, R.; Goos, K.; Güth, S.; Som, O.; Schröder, T. (2016): „Responsible Research and Innovation" als Ansatz für die Forschungs-, Technologie- und Innovationspolitik – Hintergründe und Entwicklungen. TAB Hintergrund Papier Nr. 22. Berlin (http://publica.fraunhofer.de/dokumente/N-418866.html).

Martin, B.R. (2011): The Research Excellence Framework and the 'impact agenda': Are we creating a Frankenstein monster? Research Evaluation 20 (3), 247–254 (https://academic.oup.com/rev/article/20/3/247/1560346?login=true).

Molas-Gallart, Jordi; Boni, Alejandra; Schot, Johan; Giachi, Sandro (2020): A Formative Approach to the Evaluation Of Transformative Innovation Policy. In: TIPC Working Paper (01). (http://www.tipconsortium.net/wp-content/uploads/2020/07/Paper-FORETIP-1.pdf).

Muhonen, R.; Benneworth, P.; Olmos-Peñuela, J. (2020): From productive interactions to impact pathways: Understanding the key dimensions in developing SSH research societal impact. Research Evaluation, Vol. 1, 34–47 (DOI: 10.1093/reseval/rvz003).

Peter, V.; Maier, F.; Mejlgaard, N.; Bloch, C.; Madsen, E. B.; Griessler, E.; Wuketich, M.; Meijer, I.; Wolley, R.; Lindner, R.; Bührer, S.; Jäger, A.; Tsipouri, L.; Stilgoe, J. (2018): Monitoring the evolution and benefits of responsible Research and Innovation. Summarising insights from the MoRRI project. ISBN 978-92-79-92852-9, DOI 10.2777/207020, Bericht, Luxembourg (https://op.europa.eu/en/publication-detail/-/publication/fdd7dd10-c071-11e8-9893-01aa75ed71a1/language-en/format-PDF/source-170166807).

Reale, E., Dragana Avramov, Kubra Canhial, Claire Donovan, Ramon Flecha, Poul Holm, Charles Larkin, Benedetto Lepori, Judith Mosoni-Fried, Esther Oliver, Emilia Primeri, Lidia Puigvert, Andrea Scharnhorst, Andras Schubert, Marta Soler, Sandor Soos, Teresa Sorde, Charles Travis and Rene´ Van Horik (2015): A review of literature on evaluating the scientific, social and political impact of social sciences and humanities research. Research Evaluation, 27(4), 2018, 298–308. (https://academic.oup.com/rev/article/27/4/298/3978693).

Smit, J. P.; Hessels, L. K. (2021): The production of scientific and societal value in research evaluation: A review of societal impact assessment methods. Research Evaluation. Vol 3, 323–335 (DOI: 10.1093/reseval/rvab002).

Sivertsen, G.; Meijer, I. (2020): Normal versus extraordinary societal impact: How to understand, evaluate, and improve research activities in their relations to society? Research Evaluation, Vol. 1, 66–70, DOI: 10.1093/RESEVAL/RVZ032).

Stilgoe, J.; Owen, R.; Macnagthen, P. (2013): Developing a framework for responsible innovation. In: Research Policy 42, 1568-1580. (https://www.sciencedirect.com/science/article/pii/S0048733313000930).

von Schomberg, R. (2013): A vision of responsible innovation, in: Owen, R., Bessant, J. Heintz, M. (Eds), Responsible Innovation: Opening up Dialogue and Debate. Wiley, London (DOI:10.1002/9781118551424.ch3, https://www.researchgate.net/publication/261035849_A_Vision_of_Responsible_Research_and_Innovation).

Williams, K. (2020): Playing the fields: Theorizing research impact and its assessment. Research Evaluation, Vol. 2, 191–202. (DOI: 10.1093/reseval/rvaa001).

Wittmann, F.; Yorulmaz, M.; Hufnagl, M. (2021): Impact Assessment of Mission-Oriented Policies. Challenges and overview of selected existing approaches. Karlsruhe: Fraunhofer ISI (https://www.isi.fraunhofer.de/content/dam/isi/dokumente/ccp/2021/HTF-Begleitforschung_Literaturbericht-Wirkungsmessung_2021.pdf).

AUTHORS

MERVE YORULMAZ
Fraunhofer Institute for Systems and Innovation Research (Germany)
Breslauer Str. 48, DE-76139 Karlsruhe
E: merve.yorulmaz@isi.fraunhofer.de
ORCID: 0000-0001-7810-4860

SUSANNE BÜHRER
Fraunhofer Institute for Systems and Innovation Research (Germany)
Breslauer Str. 48, DE-76139 Karlsruhe
E: susanne.buehrer@isi.fraunhofer.de
ORCID: 0000-0002-1923-1437

KEYWORDS: Responsible Research and Innovation, Evaluation Template, Scientific Impacts, Economic Impacts, Societal and Democratic Impacts, Social Labs

EVALUATION OF TRANSFORMATIONAL R&I POLICY: LESSONS LEARNED BASED ON A RETROSPECTIVE REVIEW OF FOOD SYSTEMS R&I INVESTMENT IN THE EU

SONJA SCHNEUWLY AND CAROLINE CHANDLER

DOI: 10.22163/fteval.2022.549

ABSTRACT

This paper shares our experience of developing an EU-level baseline for research and innovation (R&I) in food systems, in support of the European Commission's transformation agenda, with specific reference to the Food 2030 initiative. Food 2030 relates to the EU's mission-oriented approach to R&I, viewing it within the context of a dynamic food system with multiple dependencies and many different actors. This approach aligns with a growing recognition that, in order to achieve transformational change, the interactions and interdependencies of all components within a given system and its relationship to other systems must be considered.

In a transformative R&I system, innovation itself is no longer the end-goal but is viewed as an enabler to solve societal and environmental challenges (the end-objective). Linking such broader outcomes back to specific R&I inputs is not a straightforward endeavour. Furthermore, the inter- and transdisciplinary nature of a systems approach, as well as the nature of systems thinking itself, make it hard to define evaluative boundaries. Traditional public sector approaches to supporting R&I do not align well with such an approach, with implications for evaluating R&I policy.

The paper focuses specifically on the novel aspects of the EU's approach to framing food systems R&I and the evaluation challenges this presents, as well as how we have worked to mitigate these.

INTRODUCTION

Ensuring and creating a sustainable, climate-friendly Union is seen as a priority in order to future-proof the EU. In 2019, the Von der Leyen Commission introduced the European Green Deal – a set of policies to improve the sustainability of the European economy and ultimately achieve climate neutrality by 2050. A number of key policy initiatives further support the Green Deal's overall objectives. In relation to food, the Farm to Fork strategy aims to transform European food systems to become the global standard in sustainability while striving to supply healthy, safe, equitable, and environmentally friendly food.

Climate change and over-exploitation of planetary resources have been identified as key risks by scientists, and a particular threat to the stability of the global food system. Intensive farming practices have been linked to significant biodiversity loss and a decrease in overall soil quality, leading to reduced agricultural productivity and a reduction in the nutritional quality of food. Combined with a growing population, increased pressure on natural resources (including both land and water) and the growing prevalence of nutrient-poor diets, current production and consumption systems represent a serious threat to food and nutrition security (FNS). In this article, we therefore view food as a complex or "wicked" problem. As conceptualised by Rittel and Webber in 1973, a wicked problem is multifaceted, with no definite boundaries, and thus results in multiple different perspectives, including sometimes contradictory views regarding the main challenges, priorities and required solutions.

Research and Innovation (R&I) is considered a key tool in addressing wicked problems (including FNS). However, instead of addressing a market failure (fixing under-investment) or a systems failure (focusing on knowledge transfer and network creation), food systems R&I can be described as relating to a transformation failure. As defined by Dinges, Meyer and Brodnik (2020), the transformation-failure rationale links R&I policy to contemporary social and environmental challenges such as the Sustainable Development Goals and calls for transformative change. A transformative R&I policy response involves public and private sector actors at all levels and in all relevant sectors.

New technologies and new innovative processes are needed at all levels and in all sectors of the food value chain (including food production, processing, distribution, logistics, retail, recycling etc.) to enable such a transformation of food systems. However, a 2018 report by the SCAR Food Working Groups into R&I on food systems by European Member States identified a lack of coherence in R&I strategies and approaches to FNS at EU level and between EU Member States. The report showed that R&I activity tended to address compartmentalised elements of food supply chains rather than taking a systemic approach. Additionally, at the European level, support for food-systems R&I was distributed between different Commission services, programmes, and funding instruments.

The EU's Food 2030 policy has been designed to address this fragmentation, acting as a bridge between the European Green Deal, the Farm to Fork Strategy and the EU Framework Programmes. Food 2030 is intended to create a coherent and comprehensive approach to EU-

funded R&I on food systems transformation. It provides a roadmap demonstrating how R&I can be leveraged to ensure the long-term resilience and sustainability of the European food system in order to ensure affordable, nutritious and safe food for all European citizens within healthy planetary boundaries. It also outlines an approach for EU-funded R&I for sustainable, healthy and inclusive food to be deployed via Horizon Europe instruments, including Missions (for example the "A Soil Deal for Europe" and "Adaptation to Climate Change" missions), partnerships, and calls for proposals within Cluster 6 "Food, Bioeconomy, Natural Resources, Agriculture and Environment".

Food 2030 supports an interdisciplinary approach to food systems R&I, with the aim of strengthening policy coherence, leveraging funding and investment, supporting the development of a wide variety of innovations – from disruptive technologies to new governance processes and increasing market take-up of food products, tools and approaches and business models required to support the transition to a more sustainable and resilient EU food system. The policy framework encompasses the entire food system, taking in the whole value chain from production, packaging, transport, food environment, consumption, to waste management and health. Food 2030 identifies four priority areas for food systems R&I and ten "pathways for action", designed to provide a framework for the future-proofing of food systems through R&I action and investment (European Commission, 2020).

In late 2020, Ipsos (whose study team included the authors of this paper) was commissioned by DG RTD to carry out a comparative study related to the Research and Innovation (R&I) investment level in food systems in Europe (referred to as 'our study' throughout this paper). Our study was intended to develop a detailed understanding of the current state of play of investments in food systems R&I, both at the national and EU level, and to provide indications on the optimal level of investment that would be required to achieve the priorities identified in the future European Green Deal and Farm to Fork strategy for sustainable food systems. It mapped existing levels of public and private sector R&I investment (covering the period 2007–2020) at national and EU level against the specific priorities and pathways described within the Food 2030 strategy. By analysing historic trends for different actors within the food system, we aimed to build a preliminary view of R&I within the EU food system and identify potential future areas of intervention.

Our study involved the creation of a retrospective baseline, mapping historic levels of R&I expenditure from 2007 to 2020 against the priorities, pathways and sectors identified in the Food 2030 policy. As Food 2030 was developed after the funding being reviewed was allocated, this required an effective retrofitting of data into the specified categories. This, combined with the challenges associated with the transformative nature of the Food 2030 agenda and the wicked problem it was intended to address, led to a number of conceptual and practical challenges, which we will outline in this paper. These include considerations of how to define the scope of a systems-based approach, how to combine traditional and innovative methodological approaches to measure systemic R&I, and the implications of a systems-based approach for national and EU funding and innovation agencies.

IMPLICATIONS FOR RESEARCH AND EVALUATION: CONCEPTUAL ISSUES OF SYSTEMS TRANSFORMATION

The novel approach taken in the formulation of Food 2030 as a transformative R&I policy framework meant that the challenges associated with evaluating systems transformation applied to our study as well. This section outlines the differentiating factors of R&I policy to enable systems transformation and introduces the conceptual issues this causes for research and evaluation.

There is a growing body of literature (Zhang et al., 2018; Gill et al., 2018; Kok et al., 2019, Den Boer et al., 2021) supporting the use of a systems focus to address societal challenges linked to the global food system. R&I is considered to be a contributor to and catalyst for systems transformation. Recognising the complexity of the systems it tries to foster, as well as considering the R&I landscapes' own complexity, approaches to R&I are increasingly required to take into account the dynamism and interdependent nature of the systems they are interacting with. The systems approach is therefore more and more being applied to R&I policy. Den Boer et al. (2021) stress this notion when arguing that the complexity of food systems (implicitly applicable to systems transformation generally) requires R&I policy approaches to be both *interdisciplinary* as well as *transdisciplinary*.

Systems are interdisciplinary by nature, and beyond simply fostering multidisciplinary research, a holistic view encompassing all aspects of the system is required in order to fully understand it and thus drive change while delivering multiple as well as co-benefits and mitigating trade-offs. As defined by Den Boer et al., (2021), transdisciplinary research approaches mean that different communities of knowledge and different stakeholders (including policy makers, industry, society, SMEs) come together to "form a 'real-world laboratory' for experimentation". Complex systems mapping in cooperation with different actors is needed, as systemic transformation requires not just knowledge generation, but also appropriate implementation based on a nuanced understanding of research outputs.

A first conceptual issue which we identified when seeking to evaluate transformative R&I relates to the **lack of a measurable end objective** against which to assess impacts. Transformative R&I policy does not pursue innovation itself as the end-goal, as traditional R&I does, but rather views it as an enabler to solve societal and environmental challenges (the policy's actual objective). This means that traditional indicators such as patents, publications or market-readiness cannot in and of themselves provide evidence that R&I is indeed achieving the societal transformations for which it is being deployed. Furthermore, such indicators do not capture outcomes generated by innovation (or knowledge) systems that go beyond straightforward results. Innovation (or knowledge) systems also communicate and disseminate innovation and research outcomes to facilitate and incentive further change in the system, as well as to allow for feedback loops and evolution of knowledge (Gardeazabal et al., 2021). A shift in focus towards outcomes of systemic change (across all its levels) is therefore required (Molas-Gallart et al., 2020).

Our research addressed this issue **by using the priorities and pathways laid out in the Food 2030 policy as a common point of ref-**

erence against which to map both EU and national level investments. The pathways to action, in particular, represent the areas where the EU believes additional action and investment is required in order to achieve the high-level policy goals laid out in the EU Green Deal and the Farm to Fork Strategy. Namely, they are: 'Alternative Proteins and Dietary Shift', 'Urban Food Systems Transformation', 'The Food Safety Systems of the Future', 'The Microbiome World', 'Healthy, Sustainable and Personalised Nutrition', 'Food Waste and Resource Efficiency', 'Food Systems Africa', 'Food from the Ocean and Freshwater Sources', 'Governance and Systems Change', and 'Food Systems and Data'. By mapping R&I projects against these pathways, levels of expenditure become a proxy for innovative activity, allowing us to build up a map of hotspots, duplications, and potential gaps in achieving the EU's policy goals.

As found in our study, food-systems related R&I accounted for 3.9% of EU R&I expenditure under FP7 and Horizon 2020, amounting to **EUR 4.84 billion** in total. The share of food-systems related R&I increased sharply under Horizon 2020, reflecting the stronger mission-oriented approach taken through the newly introduced societal challenges. Over 50% of food systems funding provided through FP7 and H2020 was mapped against only four of the ten Food 2030 pathways (accounting for between EUR 550 and EUR 750 million per pathway): "Food Waste and Resource Efficiency", "Food Systems and Data", "The Food Safety System of the Future", and "Food from the Oceans and Freshwater Resources".

This contrasts with the pathways "Urban Food Systems Transformation", "Food Systems Africa", "Alternative Proteins and Dietary Shift" and "The Microbiome World", which each received between 3% and 4% (between EUR 135 and 175 million) of all relevant funding on food systems under FP7 and Horizon 2020. Furthermore, approximately 19% of projects did not fit within any of the definitions assigned to the ten Food 2030 pathways, although they aligned with the broader Food 2030 priorities.

Figure 1: EU public food-system related R&I funding (EUR million, %), per Food 2030 pathway, under FP7 and Horizon 2020 separately
Source: Ipsos analysis of CORDIS data, Inner circle = Horizon 2020 Alignment with the FOOD 2030 Pathways; outer circle = FP7

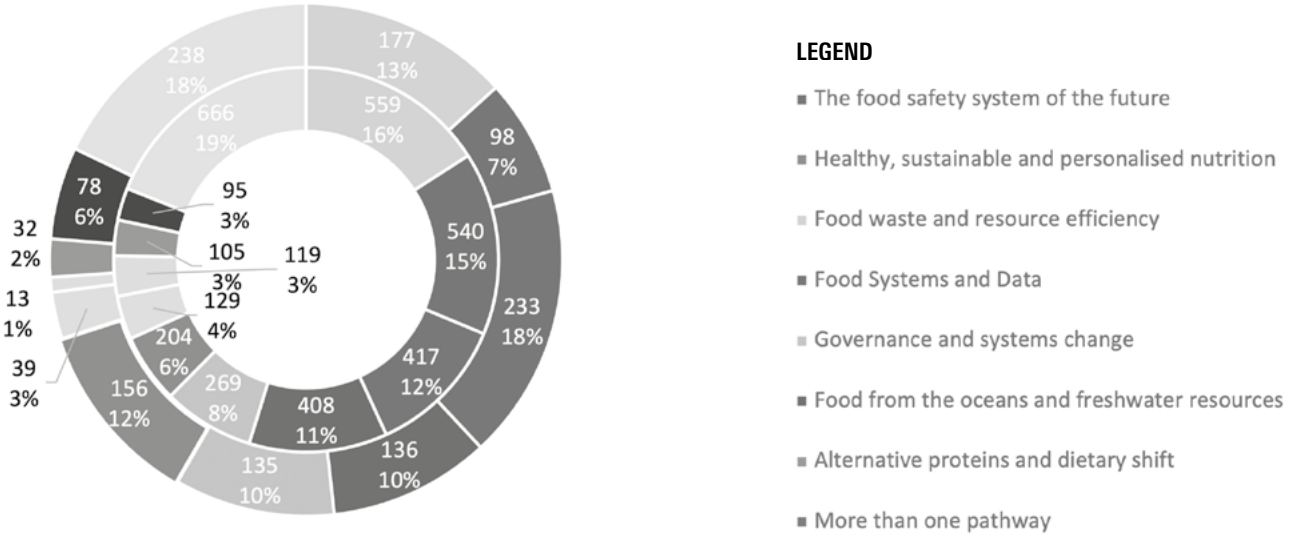
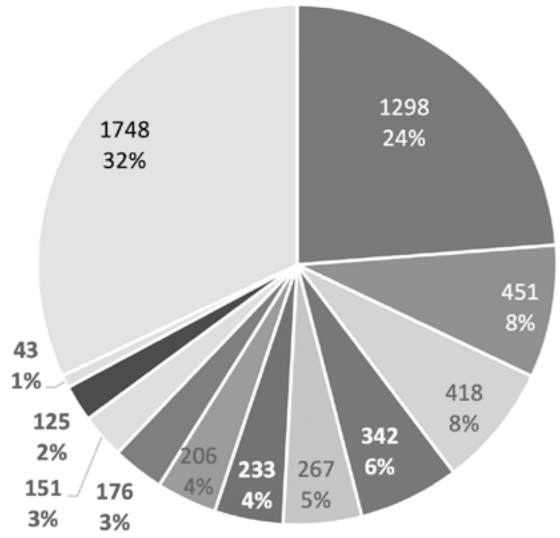


Figure 2: National food R&I expenditure by Food 2030 pathway (EUR million, %)
Source: Ipsos analysis based on 26 countries' datasets



This picture is mirrored to a large degree at the national level. Within the scope of our study, we analysed available data on food systems R&I in 26 EU Member States. Our analysis shows an estimated aggregate total of **EUR 5.5 billion** of food-related R&I spend between 2007 and 2020. While there is more even spread amongst the distribution between the ten pathways of the aggregated funding data across the 26 Member States analysed, almost one third (32%) of R&I funding did not align with any of the Food 2030 pathways.

In the private sector, this was exacerbated, with the majority of investment identified not aligning with either the Food 2030 priorities or pathways. Our study used patent data as proxy to estimate overall private sector investment in food system R&I, reaching an estimate of **EUR 93 billion** across EU Member States between 2012 and 2018[1]. Of this, 44.5% did not align with one of the four Food 2030 priorities, and almost 74% did not align with one of the Food 2030 pathways. One likely explanation for this is that private sector investments follow the individual corporate strategies of companies active in food related products and services, rather than addressing the systemic issues in food systems.

Figure 3: Patent application distribution among Food2030 Priorities, 2012-2018
Source: Ipsos analysis of Patstat data.

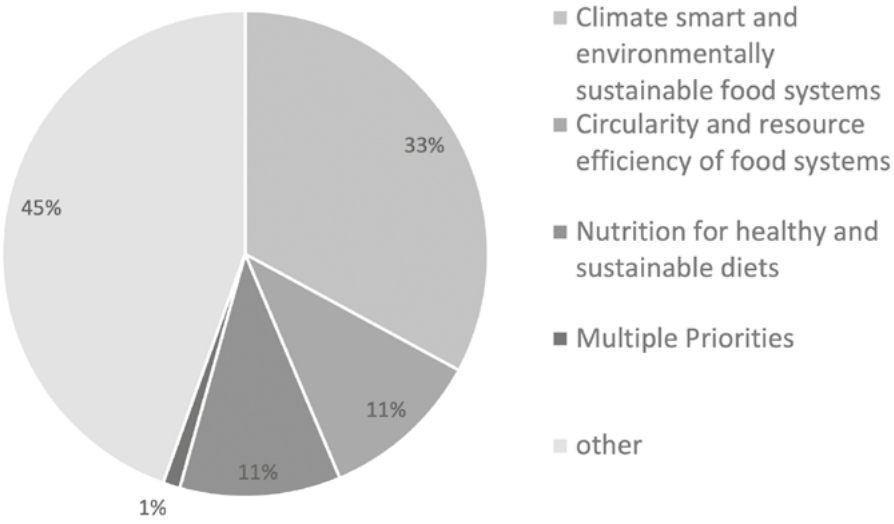
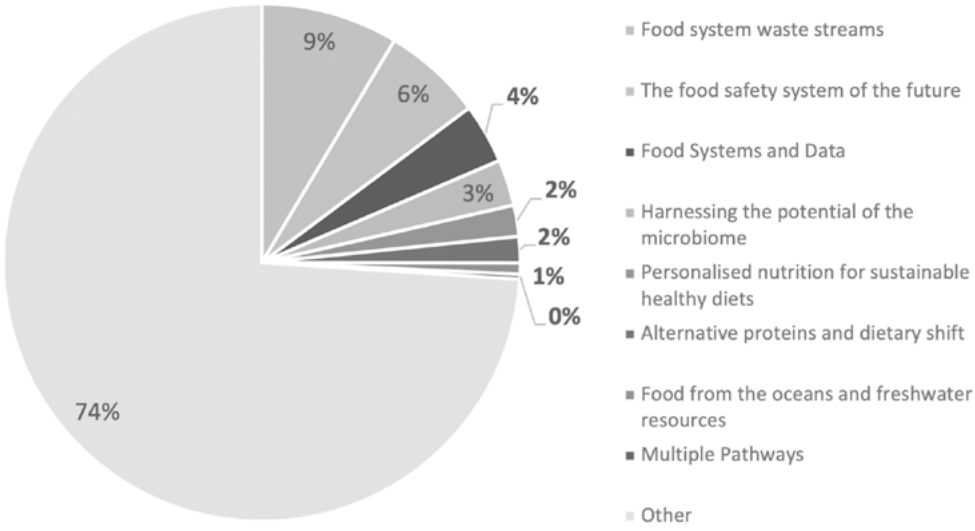


Figure 4: Patent application distribution among Food2030 Pathways, 2012-2018
Source: Ipsos analysis of Patstat data.



1 Due to data availability, the period analysed for private sector R&I spend did not fully match the period analysed for EU and national public spending

The approach taken within our study to map investments against the priorities and pathways of the Food 2030 policy in order to gain an approximation of outcomes achieved in the different areas pre-identified as pathways to change in the Food 2030 policy therefore did not prove fully satisfactory. A large share of project funding, while addressing the food system and seeking to facilitate its transformation in line with the objectives of the Food 2030 policy did not correspond to one of the ten pathways to action, but instead targeted another area. The complex nature of systems transformation means its pathways to change are numerous, and the categories used within our mapping may be too narrow to encompass these.

A second but related issue corresponds to the **non-linearity of a systems-based approach** to R&I. The nature of systems thinking, which focuses heavily on concepts of interdisciplinarity and co-creation, can make traditional approaches to attribution of impacts difficult. Interdisciplinarity "combines two or more disciplines to a new level of integration suggesting component boundaries start to break down. [It is] more than the simple addition of parts but includes the recognition that each discipline can affect the research output of the other" (ZonMW, 2020). Such influence is rarely linear but rather forges obscure causal pathways that are difficult to evaluate. This means that it can be hard to use traditional approaches to evaluation, based on identifying and assessing impacts as a series of small steps which can be used to trace the contribution of specific inputs through their conversion into activities, which in turn lead to longer-term outcomes and impacts. While the nature of our study – which is intended to map the alignment of historic expenditure with a series of policy priorities, rather than try to measure its impacts – enabled us to sidestep this issue to some extent, it remains a key point of concern when attempting to develop a comprehensive evaluation framework in order to truly understand the long-term impacts of transformational policy.

IMPLICATIONS FOR RESEARCH AND EVALUATION: PRACTICAL ISSUES OF INTERDISCIPLINARITY AND TRANS-SECTORAL PROBLEMS

In addition to the wider conceptual issues outlined above, the interdisciplinarity and trans-sectoral nature of systems transformation led to several practical issues in our study.

Candel and Pereira (2017) introduce a number of challenges related to the inter- and trans-sectoral nature of food systems policy. These include three issues of particular relevance to our mapping of R&I investment levels in food systems in Europe. The first of these relates to the breadth of sectors and stakeholders implicated in a wicked problem, and the differing priorities and interests they are likely to pursue. In the case of FNS, for example, environmental concerns may be perceived as contradictory to economic interests, which may in turn be viewed by some as undermining public health concerns. Candel and Pereira therefore highlight the need to create a shared understanding of the problem as a foundation for a resonating policy framework as a first key challenge. A second and related challenge is the need to formulate coherent policy goals that have buy-in from the different factions and design a consistent mix of policy instruments to achieve this. The vested interests of different stakeholders and sectors involved provide barriers to achieving this. The more sectors are involved, the more stakeholders need to be brought on board to buy into the narrative of change. Finally, this applies not only horizontally (i.e., between different sectors involved) but also vertically: buy-in needs to be generated from top to bottom of organisations. In the case of the EU, this would require not just the Commission to pursue objectives related to societal challenges, but also the Member States and their regions, down to the business sector and society.

Within the context of our study, this resulted in the issue faced by the study team that the inter- and transdisciplinary nature of a systems approach, as well as the nature of systems thinking itself, make it **hard to define evaluative boundaries**. The systems approach implies a broad-based and inclusive interpretation of "food systems R&I". This problem was amplified within our study by the trans-national scope of our data collection, which involved national level research in all EU Member States, meaning that any definition suffered from the potential to be "lost in translation". This posed problems when defining what was in and out of scope for our study. The approach used to identify relevant data sources therefore involved the use of two very broadly defined inclusion criteria, in conjunction with a series of pre-agreed keywords, which were further refined throughout the course of the study.

A review of national research and innovation strategies in the 27 EU Member States shows that most have **embedded R&I ambitions related to the food sector in their latest national innovation strategies**, either as a specific policy goal (Austria, Bulgaria, Germany, Ireland) or as part of a broader ambition to address societal challenges through a transformation of the food sector, often alongside the biodiversity and forestry sectors (Latvia, the Netherlands, Poland, Portugal, Romania, Spain, Sweden). Some countries, such as Finland, had already included food and agricultural R&I in strategies published in 2007. Most countries, however, only introduced it as an explicit objective more recently (in strategies published after 2018) and in many cases there is no coherent food R&I policy as such. Instead, responsibility for food-systems related R&I is subsumed within distinct agendas around agriculture, sustainability, health, education and economic growth.

For the most part, national R&I systems within Europe continue to reflect traditional policy priorities and **do not align with a systems-based approach**. Public investment in food R&I is not considered from a holistic and strategic perspective, but is instead approached in a piecemeal fashion, with Ministries for Agriculture, Economy, Education and Environment pursuing their own, often overlapping (and sometimes contradictory) policy goals. Additionally, data collected on outcomes generated is intended to monitor progress towards these individual objectives. This makes it difficult to assess progress towards systemic outcomes, as on the one hand, only partial information about certain aspects contributing to systemic outcomes is gathered, and on the other hand, trade-offs, synergies and duplications are not captured. As discussed previously, we chose to mitigate this issue by categorising spending using definitions included within the Food 2030 policy, namely the priorities and pathways defined by the European Commission. However, this approach risked missing expenditure which, while aligned with the Food 2030 priorities and pathways, was not necessarily described in a manner which allowed it to be captured by the keywords used.

Additionally, the fragmentation of responsibility for food policy (and food R&I in particular) between numerous different Ministries and other

institutions created problems both in terms of identifying relevant data (with responsibility for food R&I scattered between Ministries and Agencies) and in monitoring the outcomes of R&I investments against an overarching systems-level objective (as most data captured relates explicitly to the specific political priorities of the institution in question). It was therefore necessary to adopt an iterative bottom-up approach to data collection (casting the net widely and attempting to collect data from all parts of the system) combined with a top-down approach to data analysis, using a centralised EU policy as the common point of reference for all data collected. In reality, the majority of funding data identified fell within the remit of the Ministry for Agriculture, often complemented by limited data from the Ministry for Research and Innovation and (in some cases) the Ministry for the Economy. In many cases, data was either not consistently collected, not consistently stored, or not organised in such a way as to be able to identify food projects.[2]

Finally, it is clear that in the food R&I system, both the public and the private sectors have important roles to play. However, the requirement to calculate private sector expenditure on food systems R&I raised additional difficulties. Data collection on private sector investment is very difficult, particularly at the EU level, as there is extremely limited data available, and it does not usually provide sufficient granularity to carry out mapping equivalent to that described for EU-level and national public data collection. We therefore implemented a patent-based approach to estimating the level of private sector investment, building on a method used by Pasimeni, Fiorini, Georgakaki (2019) for estimating R&I investment levels in renewable energy using fractional counts. This allowed us to estimate a unitary cost of patents, which in turn was used to infer R&I expenditure. While patents are considered the best proxy to measuring investment levels, as they are directly linked to spending, they only capture part of all R&I activity. In the agri-food sector, patents are estimated to only represent approximately 5-10% of private sector R&I expenditure. Our study took this limitation into account when estimating overall private sector R&I spend. Our approach was felt to be a pragmatic method for understanding the scale and scope of private sector R&I investment within the EU.

CONCLUSION

In this paper, FNS is argued to be a wicked problem requiring an inter- and transdisciplinary approach to achieve (food) systems transformation. The Food 2030 Strategy is presented as a transformative strategy which aims to address this issue. We have outlined the conceptual and practical challenges this has caused for a comparative study related to the Research and Innovation (R&I) investment level in food systems in Europe we have been commissioned by DG RTD to carry out, our approaches to address these and how well they have worked.

Our research involved the identification and collation of several different national and EU datasets in order to provide a baseline picture of the overall level of investment in food-systems R&I at different levels across the EU. We have highlighted the specific challenges and limitations encountered throughout the course of this study. While Food 2030 takes precisely the inter- and transdisciplinary approach deemed necessary to address a wicked problem and provides a common framework for food systems transformation with tangible impact pathways against which Member States can measure progress, it is a novel and (to date) relatively isolated framework.

Legacy R&I policy approaches and systems still persist and were in effect throughout (most) of the period our study covers. This has implications for a retrospective mapping such as ours. These reflect a significant data gap at national level, with limited attention paid to policy priorities (such as food) which do not sit neatly within the remit of one institution, as well as a failure to capture systemic outcomes beyond the traditional R&I indicators. The Food 2030 initiative is a useful point of reference in this regard, providing a common framework for food systems transformation with tangible impact pathways against which Member States can measure progress. If food systems transformation is truly to be achieved, interdisciplinary mission-oriented food R&I strategies will need to be developed at national level with accompanying M&E strategies in order to ensure that progress towards food system is effectively monitored and measured.

Our study provided a first mapping of food systems R&I investments within the EU, but given the limitations and challenges described above, it was by necessity built using data of varying quality, completeness and granularity. To overcome this, we adopted an iterative bottom-up approach to data collection (casting the net widely and attempting to collect data from all parts of the system) combined with a top-down approach to data analysis, using a centralised EU policy (Food 2030) as the common point of reference for all data collected. Although the methodological approach described here enabled us to analyse fragmented data from multiple sources against a common transformative framework, it nonetheless represents a partial picture of food systems R&I investment within Europe.

While solutions to overcome the challenges given will usually remain partial and imperfect, they nevertheless improve on the traditional, non-systemic evaluation approaches by expanding on these and widening the scope to take in more of the edges of the hard to define evaluative boundaries than before. Gaps are still left (most notably as regards 'difficult to classify' national public and especially private sector R&I), but our study provided a useful retrospective baseline that we hope can be further refined in future research.

REFERENCES

Candel, J. and Pereira, L., (2017). Towards integrated food policy: Main challenges and steps ahead. Environmental Science & Policy, 73, pp.89-92.

Den Boer, A., Kok, K., Gill, M., Breda, J., Cahill, J., Callenius, C., Caron, P., Damianova, Z., Gurinovic, M., Lähteenmäki, L., Lang, T., Sonnino, R., Verburg, G., Westhoek, H., Cesuroglu, T., Regeer, B. and Broerse, J., (2021). Research and innovation as a catalyst for food system transformation. Trends in Food Science & Technology, 107, pp.150-156.

2 For example, project titles were included but no abstracts and/or titles were given titles such as "innovation voucher" which provided little meaningful data to analyse.

Dinges, M., Meyer, S., Brodnik, C., (2020). Key Elements of Evaluation Frameworks for Transformative R&I Programmes in Europe, ftevl Journal for Research and Technology Policy Evaluation, Vol. 51, pp. 26-40. Online access: https://repository.fteval.at/544/1/fteval_Journal_51_DOI1022163fteval2020489.pdf.

European Commission (2020). Food 2030 pathways for action – Research and innovation policy as a driver for sustainable, healthy and inclusive food systems https://op.europa.eu/en/publication-detail/-/publication/86e31158-2563-11eb-9d7e-01aa75ed71a1

Gardeazabal, A., Lunt, T., Jahn, M., Verhulst, N., Hellin, J. and Govaerts, B., (2021). Knowledge management for innovation in agri-food systems: a conceptual framework. Knowledge Management Research & Practice, pp.1-13.

Gill, M., den Boer, A. C. L., Kok, K. P. W., Breda, J., Cahill, J., Callenius, C., Caron, P., Damianova, Z., Gurinovic, M. A., Lähteenmäki, L., Lang, T., Laperrière, A., Mango, C., Ryder, J. Sonnino, R., Verburg G., Westhoek. H., Regeer, B. J., Broerse, J. E. W. (2018). A systems approach to research and innovation for food system transformation. Published by FIT4FOOD2030, https://fit4food2030.eu/eu-think-tank-policy-brief/.

Kok, K.P.W.; den Boer, A.C.L.; Cesuroglu, T.; van der Meij, M.G.; de Wildt-Liesveld, R.; Regeer, B.J.; Broerse, J.E.W. (2019). "Transforming Research and Innovation for Sustainable Food Systems—A Coupled-Systems Perspective" Sustainability 11, no. 24: 7176. Online access: https://doi.org/10.3390/su11247176.

Molas-Gallart, J., Boni, A., Schot, J., Giachi, S., (2020). A Formative Approach to The Evaluation Of Transformative Innovation Policy, TIPC Working Paper, TIPCWP 2020-01. Online access: http://www.tipconsortium.net/publication/a-formative-approach-to-the-evaluation-of-transformative-innovation-policy/.

F Pasimeni, A Fiorini, A Georgakaki (2019). Assessing private R&D spending in Europe for climate change mitigation technologies via patent data. World Patent Information, 59, December 2019. https://www.sciencedirect.com/science/article/pii/S0172219019300389

Rittel, H. and Webber, M., (1973). Dilemmas in a general theory of planning. Policy Sciences, 4(2), pp.155-169.

SCAR SWG on Food Systems (2018). Assessment of Research and Innovation on Food Systems by European Member States, Brussels. Online access: https://ec.europa.eu/research/bioeconomy/pdf/publications/Assessment_of_R_and_I_on_food_systems.pdf.

Zhang, Wei, Thorn, Jessica Paula Rose, Gowdy, John et al. (32 more authors) (2018). 'Systems thinking : an approach for understanding 'eco-agri-food systems'. In: TEEBAgriFood 'Scientific and Economic Foundations' report. TEEB for Agriculture and Food 'Scientific and Economic Foundations' report . The Economics of Ecosystems and Biodiversity, Geneva.

ZonMW (2020). R&I Policy Lab Handbook. Published by FIT4FOOD2030, https://knowledgehub.fit4food2030.eu/wp-content/uploads/2020/07/Tool_Handbook-to-support-set-up-policy-lab.pdf

AUTHORS

SONJA SCHNEUWLY
Ipsos UK
3 Thomas More Square, London, E1W 1YW
E: Sonja.schneuwly@ipsos.com

CAROLINE CHANDLER
Ipsos UK
3 Thomas More Square, London, E1W 1YW
E: Caroline.chandler@ipsos.com

KEYWORDS: food systems, food value chain, Agri-Food, transformative R&I, EU Green Deal, Food 2030, SDGs

EVALUATING A CITY LAB PROCESS IN MANNHEIM'S DISTRICT NECKARSTADT-WEST: THREE MAIN CHALLENGES FOR THE EVALUATION

SARAH SEUS AND MARIA STADLER

DOI: 10.22163/fteval.2022.550

ABSTRACT

During the last few years, city labs have emerged as promising formats to address transformative change. The aim of these formats often is to create collaborative spaces in which different stakeholders can jointly experiment with novel solutions for certain problems. While city labs start to establish transdisciplinary research settings, evaluating the effects of a city lab still brings about several challenges. In this contribution, we reflect on three main challenges that emerged in the course of evaluating a city lab in Mannheim's district Neckarstadt-West. The city lab was conducted as part of the research project SONNET (Social Innovation in Energy Transitions) and aimed to encourage social innovation in energy and thereby enable local energy transition. In the context of evaluating the city lab, we identified three main challenges that were related to a) evaluating an ongoing and open process, b) external shocks (especially in the context of the COVID-19 pandemic) and c) evaluating new forms of innovation under the concept of 'social innovation'. The main achievement of this evaluation was to trace the process of a city lab and identify changes in objectives as well as the engagement of different stakeholder groups. However, an evaluation of the city lab's outcomes remains challenging due to the openness of the process. This suggests rethinking linear evaluation models in favour of co-designing evaluation criteria in the course of the city lab process.

1 INTRODUCTION

With an increasing number of people living in cities and due to the high amount of CO2 emissions produced by cities, a growing body of literature describes cities as crucial arenas to address climate change (see e.g. Frantzeskaki et al. 2018; Evans et al. 2018). While the technical and infrastructural aspects of urban transitions surely are of great importance, the awareness for the need to address the social as well as societal aspects of urban transitions is raised. As Evans et al. (2018) describe it: "Problems associated with climate change, economic underdevelopment and social inequality are essentially urban in character. And so are their solutions." Within this discussion strand, urban experimentation has emerged as a method to explore new modes of governance, breaking routines, encouraging social innovation and empowering stakeholders that were so far not included in urban change processes. This implies the widening of the understanding of transformative change needed for transitions towards sustainability: from focusing on technical innovations only to a broader understanding of innovation that also includes social innovation.

One of the main formats for urban experimentation processes are city labs. During the last few years, a growing body of literature has discussed this format under slightly different terms (e.g. 'Reallabore' in the German discourse, see e.g. Defila and Di Giulio 2018) and has also offered different definitions of the term 'city lab'. In this contribution, we understand city labs as collaborative settings that are led by city administrations but co-designed, co-created, co-monitored and co-evaluated by further stakeholders, including researchers and citizens (Dembek et al. 2020, p. 8). One main feature of city labs is that experimentation takes place in real life contexts but is shaped by settings that are locally and temporally limited. Furthermore, city labs usually pursue specific aims, like the inclusion of stakeholder groups such as citizens or initiatives representing civil society that, so far, have been mostly excluded from (technological) innovation processes.

In this paper we discuss the evaluation of the SONNET city lab[1] in Neckarstadt-West (a municipal district of the city of Mannheim, Germany). As part of the bigger EU-funded research project SONNET, the Mannheim lab aimed at developing and testing social innovation in energy (Dembek et al. 2020). Social innovation in energy (SIE) refers to all types of changes in social relations around energy production, supply, trading or consumption. Examples are among others presuming, peer-to-peer electricity exchange but also knowledge exchange in energy dialogues or gamification for energy savings (Wittmayer et al. 2020).

The SONNET research partners of the city lab were asked to conduct an outcome evaluation of the lab and assess the results of the city lab es-

1 The project SONNET has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 837498. For more information see: https://sonnet-energy.eu/

pecially regarding how the city lab's (activities) have contributed to SIE. So far, evaluations of city lab activities have often focused on assessing the processes leading to the development of the lab's activities. The emphasis of those evaluations is then on enabling and hindering factors for co-creation processes. However, the analysis of wider outcomes of city labs and especially their transformative potential is often not taken into account. By reviewing the literature on the evaluation of social innovation, Milley et al (2018) found an emphasis on developmental evaluation approaches that focus on the process of the evaluation rather than the outcomes. One reason for this is the timing of the evaluation: city lab evaluations are often done by the research team accompanying the city lab design and implementation and therefore are conducted during the lab's lifetime and in parallel to its implementation.

The paper describes the challenges that emerged in the course of the city lab evaluation and reflects on the original ambition to assess the effects of the city lab and in particular the lab's contribution to social innovation in energy.

In this regard, we formulate the research question of this article as follows: Which challenges arise in the course of evaluating a municipal experimentation process that aims to encourage social innovation in energy?

Hence, we contribute to the current discussion on how to evaluate urban experimentation and social innovation and establish a link to the current academic discourse on how to assess transformative outcomes and the discussion on new evaluation approaches for transformative policy interventions.

The paper is structured as follows: In the next chapter, we first introduce the SONNET city lab in Mannheim's district Neckarstadt-West (NSW). In order to understand the rationale of the city lab, we elaborate on the specific characteristics that shape the city of Mannheim and the neighbourhood in which the experimentation process takes place. Chapter 3 presents the focus and methodology of the evaluation. Chapter 4 discusses the main challenges the evaluation faced with regard to a) the experimental process of the lab; b) substantial changes in the external context of the lab; c) the use of the concept of social innovation. We close with a discussion on how these challenges relate to the current discussion on the evaluation of interventions with transformative ambitions and the challenges to assess transformative outcomes.

2 THE SONNET CITY LAB IN NECKARSTADT-WEST

THE CITY OF MANNHEIM AND THE NEIGHBOURHOOD NSW

The city of Mannheim is located in the south-west of Germany. With about 320,000 inhabitants, it is among the 25 largest cities in Germany. Shaped by its industrial background and heavy industry located in Mannheim, the city has committed itself to the climate protection target and will become a climate neutral city by 2050 (status as of Sept. 2021). Thereby, the city faces different tensions. One example is keeping its status as an attractive site for economic development but at the same time reducing CO2 emissions. Furthermore, Mannheim can be characterized as a heterogeneous city including well-situated homeowners with a higher awareness for sustainability related topics but a bigger CO2 footprint on the one hand and neighbourhoods that are shaped by socio-economic difficulties, migration, low education and energy poverty on the other hand.

Mannheim is a city that claims to actively drive the transition process towards sustainability. A central aspect of its activities towards sustainability is its Mission statement "Mannheim 2030" in which the city has translated the 17 UN Development Goals into a sustainability strategy. This strategy was developed in a participatory process and translates the city's aim to encourage co-creation and (citizens) participation in decision making, the design of urban spaces and the implementation of transition pathways.

In the past, the city of Mannheim implemented several urban development projects related to energy. These urban renewal projects often focused either on new municipal districts (especially reconversions of open spaces, such as former military sites) or districts with a high socio-economic status. In recent years, the city has also been actively involved in transdisciplinary research projects. As part of one transdisciplinary project called SONNET (Social Innovation in Energy Transitions), Mannheim has created a city lab in Neckarstadt-West that aimed at encouraging social innovation in energy and thereby enabled a local energy transition.

The novelty of the SONNET city lab was to choose a neighbourhood with completely different characteristics than previous urban renovation projects as an experimental space. NSW is a densely-populated district, with few green and recreational spaces. The majority of inhabitants in NSW are tenants that live in older apartment buildings, which need to be refurbished. Social deprivation, a high unemployment rate, migration and social exclusion characterise its population. In this context, so far no priority has been given to participatory projects on the topic of energy. Hence, no blueprint existed for the city lab stakeholder with regard to how to involve a densely populated inner-city district in a process of participatory energy transition.

THE CITY LAB NSW

The SONNET city lab in NSW started from a broader definition of social innovation in energy, defining SIE as changes in social relations around local energy use and consumption. This was to be achieved by eliciting new ways of communication and interaction between the city administration and local stakeholders, stimulating citizen participation and inducing shifts in roles and responsibilities of the participant stakeholder. Therefore, the city lab aimed at collaboratively designing energy-related activities and implement them during the city lab's lifetime. All activities were to be designed to address the specific needs of the Neckarstadt-West neighbourhood.

The lab was conducted between December 2019 and August 2021. It started with three design thinking workshops involving city administration staff and stakeholders working in NSW. The objectives of the design thinking workshops were to bring stakeholders together around the topic relevant to energy transition and to develop activities that could then be implemented in NSW. The first COVID-19 pandemic induced lock-down entered into force a couple of days after the last design thinking workshop and stopped further work of the workshop participants. However, this break was also a possibility to re-think the initial objectives and planning of the city lab. It had become clear during the first months of the city lab that the design-thinking process was not well suited to represent

the diversity of already existing initiatives and stakeholders in NSW and to reach out to the citizens. Instead, it was important to get to know the existing stakeholder structures of the neighbourhood. NSW as a dense inner-city neighbourhood has already a lively scene of associations, although only few initiatives on the topic of energy exist. The city lab was the possibility to see which stakeholder groups were interested in joining forces on the topic of energy transition.

Consecutively network building with stakeholders in NSW and finding multipliers and mediators for the topic had to be prioritized. The phase of network building and stakeholder engagement took place between summer and autumn 2020 and culminated in a first NSW-stakeholder group event in December 2020. It targeted especially professional (full-time professionals, e.g. teachers at schools, neighbourhood managers etc.) and organized stakeholders (volunteers, e.g. associations, local citizen networks) of NSW and gave them a new platform to brainstorm topics to be prioritised. Due to the COVID-19 pandemic, the discussion on the development of activities continued within a narrow group of people.

While initially the city lab aimed for co-designing concrete activities, the final choices for the two show-case activities was made by the core actors involved in the city lab in spring 2021. These changes to the initial plan were a result of the readjustment processes of the lab and reinforced by the restrictions imposed by the COVID-19 pandemic. From a practical perspective, the activities were chosen because they both could be implemented despite the retractions that existed due to the COVID-19 pandemic. However, these activities had never been implemented in Mannheim before and reinforced the aim of the city lab to experiment with new methods for engaging citizens and local stakeholders in a dialogue process. In line with the aim to include local stakeholders in a broader transition process, that activities focused on knowledge exchange and awareness raising and allowed to inform and involve different stakeholder groups. The two activities - 1) the Mobile Green Room® and 2) the KliMAthon app - were both implemented in summer 2021.

- The Mobile Green Room® is a planted, container-like platform that can be transported and therefore allows to temporarily display urban green in densely built environments.[2] It serves as a prototype for greening the urban areas and allows citizens to experience the advantages of urban green. In Neckarstadt-West it was also used as a platform for outdoor events or informal meetings as it provides space to sit or stand on it. As part of the city lab in Neckarstadt-West, a Mobile Green Room was installed for 12 weeks in different locations in the district and used by different stakeholders, such as a local church community, a school or the neighbourhood management for activities.
- The KliMAthon[3] is an app-based competition that encourages participants to save CO2 emissions. It allows to calculate a personal CO2 footprint, provides tips for climate friendly behaviour and encourages to participate in 'challenges' such as taking the bike to work or abstain from dairy products for a certain period of time. For 42 days, citizens were invited to use the app and participate in challenges in order to save emissions together and create awareness for sustainable behaviour.

The selection of the two activities illustrates a broadening of the scope of the city lab activities beyond energy related topics towards sustainability in a more general way. The city lab participants regarded this broadening of the thematic scope as necessary in order to better reach out to the local Neckarstadt-West stakeholders. The original narrow focus on 'energy only' was described as being too abstract to attract the attention of locals, however, the focus on sustainability - especially on enhanced living conditions - was more in line with the citizens´ needs.

The next section discusses the challenges that arose during the evaluation process.

3 EVALUATION FOCUS AND METHODOLOGY USED

In order to assess the effects of the city lab and especially changes in social relations in the field of energy in NSW, the evaluation chose the following evaluation criteria and formulated the evaluation questions displayed in the table.

Eval. Criteria	Explanation
Relevance	The relevance of the activities with regard to the needs of the inhabitants of the neighbourhood Neckarstadt-West.
Coherence	The coherence and embeddedness of activities with the Mannheim strategy "Mannheim on Climate Track" and the embeddedness of various non-SONNET related activities existing in Neckarstadt-West.
Inclusiveness	The inclusiveness of the process of the city lab: this aspect was twofold and looked into a) whether the relevant stakeholders in Neckarstadt-West had been included in the city lab and b) whether professional stakeholders from outside Neckarstadt-West, especially staff from different city departments had been involved in the city lab and how.
Effectiveness	The effectiveness of the lab with a specific focus on how the city lab had contributed to changing social relations (in particular in the energy field), for example new networks of actors, changes in communication patterns or even new organisation structures (social innovations).

Table 1: Evaluation Criteria of the SONNET City Lab Evaluation

2 https://www.mannheim-gemeinsam-gestalten.de/dialog/informationen/mobiles-gruenes-zimmerr-der-neckarstadt
3 The App was developed and provided by worldwatchers GmbH: https://www.worldwatchers.org/

While the evaluation criteria were suggested by the SONNET project, the exact formulation of the evaluation questions and their translation into descriptors was re-adjusted during the evaluation process. This allowed reacting to changes in the city lab process.

From a methodology point of view the evaluation was inspired by the method of process tracing (George and Bennett 2005). For the data collection and analysis we concentrated on qualitative methods as they allowed greater flexibility to adapt to the processual character of the lab and trace the development of changes. The evaluation questions were translated into qualitative descriptors that qualified the degree of changes occurred during the city lab implementation.

Our main data collection methods were interviews and participatory observations. At the heart of the evaluation were interviews with stakeholders involved in the two show-case activities, the Mobile Green Room® and the KliMAthon app. Overall, 10 interviews were conducted between May and the beginning of August 2021. The data was complemented by interviews with Mannheim stakeholders not primarily related to the city lab activities in NSW and observations at different events conducted during the city lab, such as the design thinking workshop or stakeholder events.

During the course of the evaluation, it became clear that the focus of the evaluation had to be put on the implementation process of the city lab. The evaluation of the outcomes of the city lab, however, could not be realised as originally intended. Especially the assessment of SIE development and its transformative potential remained rudimentary. The next chapter discusses the evaluation's challenges that finally led to the shift in the evaluation focus.

4 DISCUSSION OF CHALLENGES FACED DURING THE EVALUATION

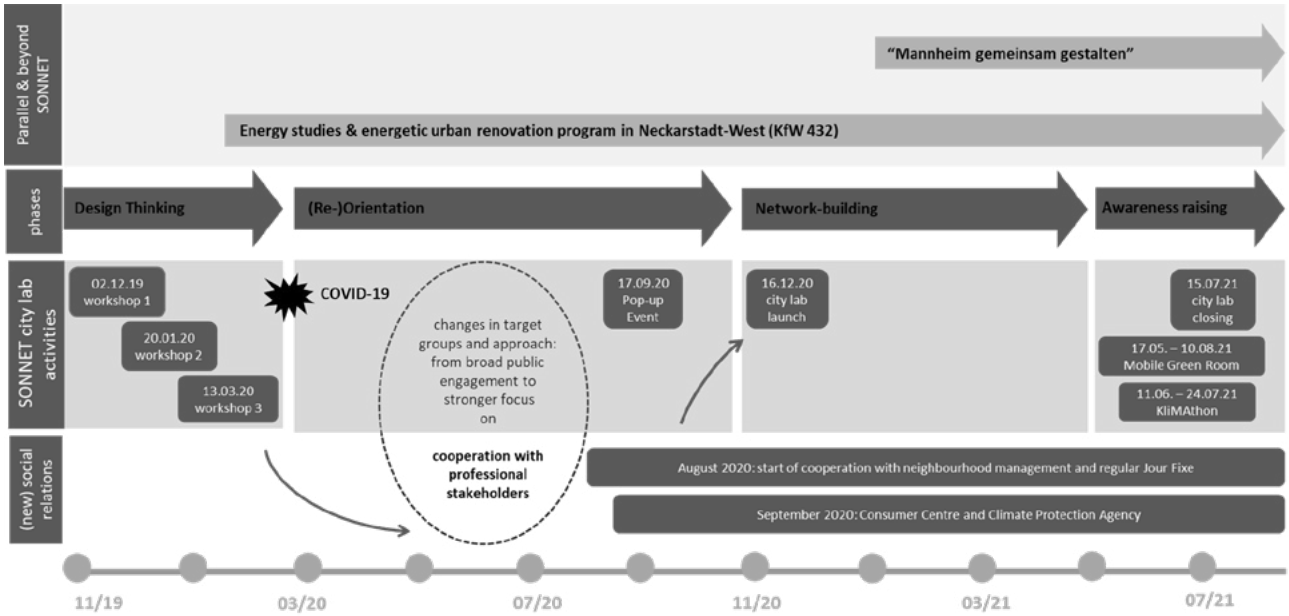
4.1 CHALLENGES RELATED TO EVALUATING AN ONGOING EXPERIMENTAL PROCESS

As described in chap 1.2., the SONNET city lab was characterised by changes in methods, activities, target groups and the reflection on objectives. We argue that this process of permanent readjustments is characteristic for an experimental process. In the literature on (urban) experimentation the experimental processes show four phases: a) launch of the process, b) preparation phase (especially choosing the experimental activity, the methods to implement, the indicator to measure it, and the stakeholder involved; c) the implementation phase of the experiment; d) evaluation / reflection phase (e.g. Knieling et al. 2021). This is also how SONNET foresees the city lab implementation (Dembek et al. 2020).

For the evaluation of the city lab, two challenges arose: First, the assessment of the lab's contribution to perceived change was difficult due to the fluid boundaries between the lab and other parallel activities. Second, the readjustments with regard to goals and target groups during the lab implementation process asked for a flexible approach to the evaluation especially regarding the assessment of goal attainment.

In order to address these challenges, the evaluation team traced the implementation process backwards. The starting point was the reconstruction of the overall city lab process at the end of the lab's lifetime (summer 2021). The next figure shows the reconstructed process. In retrospective, four phases of the city lab have been defined which have involved changes in target groups and subsequent adaptation of activities, events and tools.

Figure 1: Reconstructed process of the SONNET city lab
Source: own illustration, Fraunhofer ISI.



In the SONNET city lab NSW the four phases mentioned in the literature could not be clearly distinguished. The "experiment" started already with the choice of the neighbourhood NSW, as no blueprint existed in Mannheim how to encourage social innovation related to energy in a neighbourhood similar to NSW. The city lab's main activity (and its main achievement) was to connect stakeholders who were willing to join forces on the topic of energy transition in the district. In this sense, the overall city lab implementation was an open process with regard to the stakeholder groups involved, the choice of activities and to some extent the objectives of the lab.

With regard to methods and tools, the city lab was a possibility to test out which methods would work well, under which preconditions and for which types of stakeholders (e.g. professional stakeholders, initiatives and organisations of NSW, citizens of NSW). Testing methods and tools for stakeholder involvement was an explicit aim of the city lab from the beginning but was intensified during the process of the lab. As the organizers of the SONNET city lab were also involved in other projects in the neighbourhood (e.g. a program for energetic refurbishment) and activities on the city level (e.g. the Mannheim strategy or citizen involvement process "Gemeinsam gestalten") the SONNET city lab eventually fitted well into the overall city's projects portfolio on sustainability transition.

The adjustments during the lab implementation also implied shifts in the objectives of the city lab. The original aim to design and implement activities related to energy transition in the city lab's lifetime and with participation of local stakeholders, especially its citizens, had to be revised. The evaluation acknowledged these changes and focused on assessing the network building and interaction processes of stakeholders around energy related topics. Furthermore, it took up the question which participation methods were suitable for different stakeholder groups in order to induce changes or create new social relations.

The evaluation took place during the final phase of the city lab, i.e. still during the lifetime of the experiment. However, due to the short lifetime of the overall lab (1,5 years including a longer stand-still period caused by the COVID pandemic) it was difficult to assess changes in social relations, especially with regards to their innovative potential (i.e. mainstreaming potential and sustainable application).

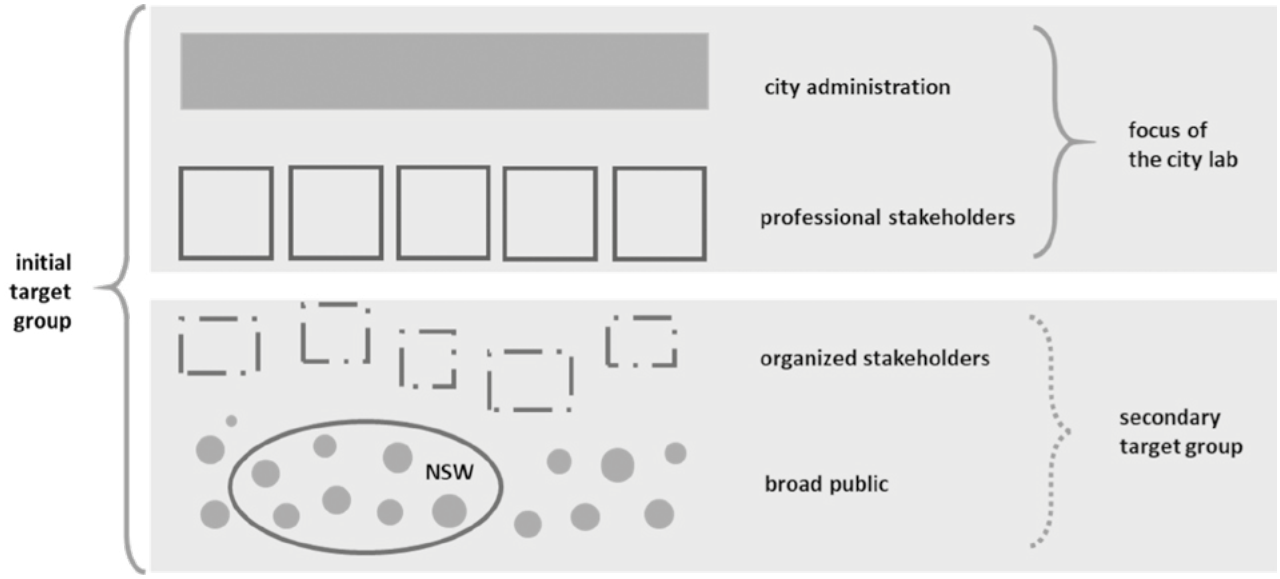
4.2 CHALLENGES RELATED TO EXTERNAL SHOCKS

The COVID pandemic started four months after the launch of the city lab. The first lockdown in spring 2020 destroyed the initial plan to gather (physically) the local initiatives and contact citizens in the NSW neighbourhood and discuss activities related to energy transition. Due to this external shock, the priorities of local NSW stakeholders changed considerably. Health issues dominated the local political agenda to the disadvantage of the already non-prioritised topic of sustainability in general and energy transition in particular. The city lab continued its work, however, with a small group of core participants, all of them "professional stakeholders", i.e. people working for the neighbourhood as part of their paid job, such as the staff of the city administration. On the other hand, volunteers of local initiatives but also school staff had only limited possibilities to participate in the discussions around the city lab, especially in the times of the COVID lock-downs.

The next figure presents the shifts in target groups during the lab implementation.

The COVID pandemic affected the evaluation insofar as only few on-site visits could be realised, limiting the possibility for direct observation

Figure 2: Target groups of the SONNET city lab
Source: own illustration, Fraunhofer ISI.



e.g. in meetings or a newly created exchange (regular meeting of the stakeholders involved to discuss the local energetic renovation process in NSW) or spontaneous exchanges between the team of evaluators and the city lab participants.

4.3 CHALLENGES RELATED TO THE VAGUE CONCEPT OF SOCIAL INNOVATION

With the overall aim of the SONNET city lab in NSW to encourage social innovation in energy (SIE), one of the main challenges faced in the evaluation process was that a plurality of definitions of social innovation exists among researchers as well as practitioners.

In academic discourse, differences in defining social innovation (SI) exist among scholars who follow a normative definition that highlights the role of social innovation as 'good for society' (Murray et al. 2010) and those using it as an analytical definition interested in tracing the development of social innovation (Howaldt et al. 2015). Also the subject of SI is divers. It can refer to changes in social relations (Avelino and Wittmayer 2017), novel practices, e.g. related to sustainable consumption (Jaeger-Erben et al. 2017) or novel business models, e.g. contributing to energy justice (Hiteva and Sovacool 2017). One central aspect of social innovation, however, is the empowerment of social groups that so far have been excluded from participating in innovation processes.

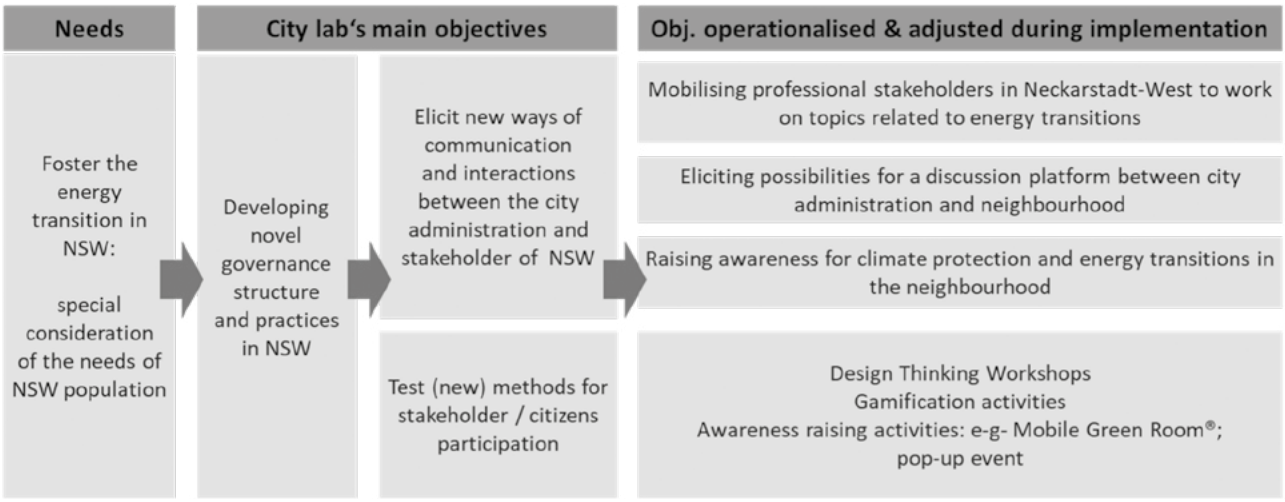
For the SONNET project, the main interest was in understanding how social relations around energy are changing and what conditions enable or impede the transition towards a more sustainable energy system. In this sense, SONNET defines SIE as "(combinations of) ideas, objects and/or actions that change social relations and involve new ways of doing, thinking and/or organising energy" (Wittmayer et al. 2020). The SONNET city lab in NSW embedded this definition of SI in its overall aim, to *"develop novel urban governance structures and practices for enabling social innovation in the energy sector."*

However, this definition left room for interpretation with regard to what social innovation in energy might be. Consequently, the city lab stakeholders rarely used the concept of social innovation. Especially in discussions with citizens and local initiatives, the city lab responsible translated SIE as was "social aspects of energy". SONNET's research on the role of policy making for social innovation showed that the awareness for social innovation among policy makers still needs to be increased and a shared understanding of SI developed (Rogge et al. forthcoming). This also applies for the local level: While the awareness for the important role of social innovation starts to increase among policy makers on the local level, a shared understanding of SI and a way to implement it is still missing.

The evaluation faced the challenge how to work with such a fuzzy definition as two options for analysing SIE seemed possible: SIE could both be seen as a 'means' in the city lab process (i.e. such as in the SONNET definition of SIE as "changing social relation") or rather as a 'result' of the city lab, (i.e. as suggested by the city lab objective of "enabling social innovation in energy").

One task of the evaluation was to operationalise the overall aim of the city lab into distinct and clear subordinated objectives but also to name the broad rational of the lab ("foster energy transition in NSW"). The next figure presents the final version of the hierarchy of objectives, which was elaborated at the very end of the evaluation process. On purpose, SIE is not explicitly mentioned here, as their development is considered as an outcome of the city lab.

Figure 3: Objectives of the SONNET city lab
Source: own illustration, Fraunhofer ISI.



As a consequence of the unclear definition of SIE, two different evaluation designs were possible: analysing SIE as an activity of the city lab or tracing the process that produces SIE. The evaluation started by evaluating SIE as an activity, namely the two show-case activities implemented in summer 2021. However, it became clear during the evaluation that focusing on such a narrow part of the overall city lab would not have been sufficient to capture the complexity of the city lab. Thus, after the first interviews, we shifted the focus in order to take the overall development process of the city lab into account. Rather than defining ex-ante activities as social innovation and assess their effects on the city lab, we understood SIE as a possible result of the city lab. In line with the definition of SIE as "changes in social relations" (in the context of energy), we focused our analysis on the identification and interactions of stakeholder groups as well as the tools and methods that structured exchange and participation. In the particular case of the SONNET city lab, the following aspects proved to be important for encouraging SIE:

a) Mapping stakeholders before engaging in a participatory design process;
b) Reflecting on suitable participation formats for different stakeholder groups;
c) Establishing a new communication process between stakeholders who had formerly not interacted;
d) Taking into account the needs and external contextual constraints of the involved stakeholders.

In this sense, the SONNET city lab allowed to gain knowledge on enabling and impeding conditions for SIE in the NSW neighbourhood. At this point in time we cannot tell whether the new stakeholder configuration and interaction practices will be continued in the NSW or even inspire processes in other districts of Mannheim. One has to acknowledge that the city lab was conducted over a short period only – but changes in social relations are processes that require time to develop and become institutionalized (Hielscher et al. 2020). However, the evaluation concluded that the city lab has successfully kicked-off a stakeholder identification and mobilization process on energy topics in NSW. This process is likely to be continued after the city lab has closed, as the neighbourhood is currently participating in a five-year urban renovation programme (funded by the German KfW-Bank). In this sense, the city lab can be understood as an important first step in a longer urban transition process.

5 DISCUSSION AND OUTLOOK

In this article, we reflect on the challenges faced during the evaluation of the SONNET city lab process in Mannheim's district Neckarstadt-West. The challenges that the evaluation was confronted with arose from its main features that were a) the experimental character of the project and b) the concept of social innovation as a central conceptual framework of the project. Furthermore, the changes in planning of the experiments as a result of the COVID pandemic added another challenge.

Recent literature on the evaluation of transformative innovation policies discuss different evaluation design features that could be useful to assess these specific types of policy interventions. Molas-Gallart et al. (2020) present "six guiding principles for transformative innovative policy evaluation". One of the principles is the idea to use a flexible theory of change that is readjusted during the evaluation. We found this idea very useful with regard to the reorientation of objectives and activities of the SONNET city lab. Indeed, we readjusted the logic chain leading to expected outcomes several times in order to be in line with the actual activities and rationales of the stakeholders, but also following a deeper understanding of the city lab's experimental process.

In the evaluation of the city lab NSW, not all of the six principles could be implemented, especially the claim to conduct a formative evaluation with the active participation of stakeholders (affected by the evaluation) could not be realised. However, we acknowledge that a stronger involvement of at least the main stakeholders of the city lab (from the city administration and the neighbourhood) would have allowed going deeper in assessing the effects of the city lab instead of focusing mainly on the process and conditions for implementation.

The other side of the claim of a more inclusive and participatory mode of evaluation, is the (new) role for the evaluators. In the SONNET city lab the role of the evaluator was not clearly defined. The SONNET project assigned to the research partner (and evaluator) a role as an involved partner in the city lab, e.g. as facilitator, mediator or as a partner for critical reflection. In the SONNET city lab the evaluator's role should be rather described as an external assessor. Exchanging the ideal roles (and their evolution) of each party should be an integrated part of this new type of "co-productive" evaluation.

The period in which the evaluation had to take place - namely during the lifetime of the city lab - hindered looking at outcomes. This seems to be a general problem of transdisciplinary social labs. Generally, these labs have a dedicated research team that monitors the implementation process scientifically and compares different settings (e.g. different city labs such as in the SONNET project). However, the mandate of these researchers ends with the implementation of the city labs. To our knowledge, ex-post outcome evaluations of city labs are seldom conducted.

In our case it proved very helpful to focus on the evaluation criteria of 'relevance', inclusiveness' and 'coherence' and to put the emphasis of the analysis on the design of the interventions as a crucial factor that influences the effective implementation and development of effects. (Mickwitz et al. 2021)

Ghosh et al. (2020) discuss the need for new outcome categories (especially complementing the traditional STI outcome categories and indicators), the so-called "transformative outcome". Social innovation can be understood as such a new type of outcome. Research on the role of policy making for social innovation suggests that more research is needed on the possible impact pathways of SI and the development of indicators along the pathways (Rogge et al. forthcoming). While on the EU and the national level policy strategies for social innovation are emerging, the concept of social innovation needs to be broken down to the local context and the specific aim of the experimental process.

Our contribution highlights the challenges that evaluations of experimental and transdisciplinary policy measures are confronted with. Current discussions in the evaluation community provide interesting approaches that could be further explored and tested in future evaluations.

6 BIBLIOGRAPHY

PUBLICATION BIBLIOGRAPHY

Avelino, Flor; Wittmayer, Julia M. (2017): Transformative social innovation and its multi-actor nature (Atlas of Social Innovation).

Defila, Rico; Di Giulio, Antonietta (2018): Reallabore als Quelle für die Methodik transdisziplinär und transformativen Forschens - eine Einführung. In Rico Defila, Antonietta Di Giulio (Eds.): Transdisziplinär und transformativ forschen. Eine Methodensammlung. Wiesbaden, Germany: Springer VS (Open), 9-35.

Dembek, Agata; Danokowska Alicja; Struminska-Kutra, Marta (2020): SONNET - SOCIAL INNOVATION IN ENERGY TRANSITIONS Co-creating a rich understanding of the diversity, processes, contributions, success and future potentials of social innovation in the energy sector. D4.1: Report on transdisciplinary research protocol for six co-creating SIE city labs, checked on 11/2/2021.

Evans, James; Karvonen, Andrew; Raven, Rob (2018): The experimental city. New modes and prospects of urban transformation. In James Evans, Andrew Karvonen, Rob Raven (Eds.): The experimental city. First published in paperback. London and New York: Routledge (Routledge research in sustainable urbanism).

Frantzeskaki, Niki; Hölscher, Katharina; Wittmayer, Julia M.; Avelino, Flor; Bach, Matthew (2018): Transition Management in and for Cities: Introducting a New Governance Approach to Address Urban Challanges. In Niki Frantzeskaki, Katharina Hölscher, Matthew Bach, Flor Avelino (Eds.): Co-creating Sustainable Urban Futures. A Primer on Applying Transition Management in Cities. Cham: Springer (Future City Ser, v.11), pp. 1–40.

George, Alexander; Bennett, Andrew (2005): Case studies and theory development in the social sciences. Cambridge, Massachusetts: MIT Press.

Ghosh, Bipashyee; Kivimaa, Paula; Ramirez, Matias; Schot, Johan; Torrens, Jonas (2020): Transformative Outcomes: Assessing and Reorienting Experimentation with Transformative Innovation Policy Transformative Outcomes. In TIPC Working Paper, TIPCWP (02). Available online at https://www.tipconsortium.net/publication/transformative-outcomes-assessing-and-reorienting-experimentation-with-transformative-innovation-policy/, checked on 9/9/2021.

Hielscher, Sabine; Wittmayer, Julia M.; Durrant, Rachael (2020): SONNET's Methodological guidelines for case study analysis.

Hiteva, Ralitsa; Sovacool, Benjamin (2017): Harnessing social innovation for energy justice: A business model perspective. In Energy Policy 107 (March), pp. 631–639. DOI: 10.1016/j.enpol.2017.03.056.

Howaldt, Jürgen; Kopp, Ralf; Schwarz, Michael (2015): On the theory of social innovations: Tarde's neglected contribution to the development of a sociological innovation theory. Weinheim: Beltz Juventa.

Jaeger-Erben, Melanie; Rückert-John, Jana; Schäfer, Martina (2017): Soziale Innovationen für nachhaltigen Konsum. Wiesbaden: Springer Fachmedien Wiesbaden.

Knieling, Jörg; Kretschmann, Nancy; Nell, Rebecca; Pfau-WEller, Natalie (2021): Wissensproduktion und Lernprozesse am Beispiel von Reallaboren zur Klimafolgenanpassung in Halle (Saale) und Mannheim. Real-world laboratories in theory and practice: Reflection of the research design regarding the sustainable transformation of urban spaces. In Raumforschung und Raumordnung | Spatial Research and Planning 79 (4), pp. 438–452. Available online at http://publica.fraunhofer.de/documents/N-640351.html, checked on 11/2/2021.

Mickwitz, Per; Neij, Lena; Johansson, Maria; Benner, Mats; Sandin, Sofie (2021): A theory-based approach to evaluations intended to inform transitions toward sustainability. In Evaluation 27 (3), pp. 281–306. DOI: 10.1177/1356389021997855.

Milley, Peter; Szijarto, Barbara; Svensson, Kate; Cousins, J. Bradley (2018): The evaluation of social innovation: A review and integration of the current empirical knowledge base. In Evaluation 24 (2), pp. 237–258. DOI: 10.1177/1356389018763242.

Molas-Gallart, Jordi; Boni, Alejandra; Schot, Johan; Giachi, Sandro (2020): A Formative Approach to the Evaluation Of Transformative Innovation Policy. In TIPC Working Paper (01), checked on 2/23/2021.

Murray, Robin; Caulier-Grice, Julie; Mulgan, Geoff (2010): The open book of social innovation. London: Young Foundation; National Endowment for Science, Technology and the Arts (Social Innovator Series : ways to design, develop and grow social innovation).

Rogge et al (forthcoming): Co-creating strategies for navigating multi-level policy dynamics to encourage SIE – reflections. SONNET project deliverable D2.4

Wittmayer, Julia M.; Hielscher, Sabine; Rogge, Karoline S.; Avelino, Flor (2020): Report on SONNET's initial conceptual framework.

AUTHORS

SARAH SEUS*
Fraunhofer Institute for Systems and Innovation Research ISI
Breslauer Str. 48, 76139 Karlsruhe, Germany
E: sarah.seus@isi.fraunhofer.de
phone: +49-721-6809-344
ORCID: 0000-0003-0791-5919

MARIA STADLER
Fraunhofer Institute for Systems and Innovation Research (Germany)
Breslauer Str. 48, 76139 Karlsruhe, Germany
Technical University of Darmstadt, Institute of Sociology, Germany
E: Maria.Stadler@isi.fraunhofer.de
ORCID: 0000-0001-7114-3103

*corresponding author

PATIENT AND PUBLIC INVOLVEMENT AND ENGAGEMENT (PPIE): FUNDING, FACILITATING AND EVALUATING PARTICIPATORY RESEARCH APPROACHES IN AUSTRIA

RAPHAELA ELISABETH KAISLER AND THOMAS PALFINGER

DOI: 10.22163/fteval.2022.551

ABSTRACT

The LBG OIS Center established a new Patient and Public Involvement and Engagement (PPIE) Implementation program aiming at 'active involving' public members in research across different phases of the research cycle – from setting the agenda to disseminating results – and its governance. The program offers funding and facilitation of these PPIE activities. The first PPIE pilot call was launched in Autumn 2020. It supports researchers in Austria with up to EUR 60.000 in order to implement their PPIE activities. In addition, the program offers support in the form of consultation, training, knowledge exchange and networking opportunities. One important characteristic of the selection process is the composition of the expert panel, bringing together transdisciplinary expertise from different areas (scientific experts, patients, and students). The expert panel recommended 11 out of 25 PPIE projects for funding (success rate 44%). 45% of the applicants participated in the support offers prior to the call and 52% in the continuing support offer after the call had been closed. Based on our online surveys, overall, participants were very satisfied with the support offers. Learnings of the first call address the eligibility of applicants. In the selection meeting, we found that different understandings of 'active involvement' were negotiated among experts. However, this was not a problem due to the open and collaborative atmosphere and mutual learning opportunity for experts. The panel suggested opening the call to non-research bodies, which indicates small changes in the application format – e.g. video and text-based applications in German and English. Despite of small adaptions in the second PPIE Pilot Call 2021, it seems that the funding instrument was appropriate and reflects a low-threshold offering for researchers introducing public involvement activities in their work.

BACKGROUND

Patient and public involvement and engagement (PPIE) in research is an important driver for societal impact of science and its capability to develop new solutions for existing challenges (Greenhalgh et al. 2019). One of the main drivers is the United Kingdom, which introduced PPIE in the national research agenda and research funding (National Institut of Health Research 2021): research carried out 'with' or 'by' members of the public rather than 'to', 'about' or 'for' them (Hayes, Buckland, and Tarpey 2012). Such approaches to engaging the public in research have increasingly come into the focus of national and international policy actors often framed as citizen science, public engagement, and public involvement in policy documents (Hecker et al. 2019; Bundesministerium für Verkehr 2016; Bundesministerium für Wissenschaft 2015; European Commission 2014). Hecker et al. (2019) explored the conceptualization of citizen science and found that policy documents attribute educational benefits to citizen science by fostering scientific literacy, individual learning, and skill development, as well as by facilitating environmental stewardship.

In a recent analysis of 34 reviews, Ocloo and colleagues (2021) explored barriers and enablers of patient and public involvement in health-related research. The authors identified adequate funding and resources, the lack of training opportunities for the public and professionals, and the lack of general support structures, such as emotional, practical and financial support as key barriers among others. Health researchers also emphasized the emotional component of working with PPIE, which is both rewarding and burdensome, and requires practical as well as social support (Boylan et al. 2019). Further, researchers' positive attitude and experience towards PPIE are key factors in the successful implementation of public involvement activities in research (Boylan et al. 2019, Nathan et al. 2006, Thompon et al. 2009, Ocloo et al. 2021).

The situation is similar in Austria. Here, too, there is a gap in the implementation of such participatory approaches among researchers and a lack of instruments for funding such approaches in the Austrian research landscape. Challenges of implementing PPIE practices address the lack of awareness and knowledge about the PPIE concept in the local scientific communities, the lack of appreciation of the value of involving patients as 'experts by experience' and fear of violating research ethics if PPIE activities are carried out without formal ethical approval (Kaisler et al. 2021). To overcome these challenges, the Ludwig Boltzmann Ge-

sellschaft (LBG) Open Innovation in Science (OIS) Center established the PPIE Implementation Program in 2020. It intends to fill the gap and institutionalize support for public involvement activities in Austria. The PPIE Implementation Program is funded by the National Research Foundation for Technology and Development. The PPIE Implementation Program is embedded in a wider 'open innovation in science' framework fostering collaboration among different stakeholder groups to enable transdisciplinary collaboration. This requires an open mind-set and open research practice which allows for thinking beyond the research discipline and academic framework in order to generate scientific insights and translate them into innovations (Beck et al. 2020).

CO-DEVELOPMENT OF THE PPIE IMPLEMENTATION PROGRAM

To systematically introduce public involvement at LBG and in Austrian research institutions, we co-developed the PPIE 'How to' Guide for Researchers (Kaisler & Missbach 2019) aiming to support researchers in implementing public involvement activities in their research. In a series of five co-creative workshops, citizens, patient advocates and LBG researchers co-created principles of PPIE in research and project steering structures, self-assessment checklists, and monitoring of PPIE activities in research projects (Kaisler & Missbach 2020). In the last workshop, we discussed potential funding models with all stakeholder groups based on the PPIE 'How to' Guide for Researchers. The discussion covered three topics with the aim of co-creating action plans to establish a public involvement focus at LBG:

1. funding structures to implement public involvement activities,
2. support structures to facilitate implementation,
3. and the evaluation of public involvement activities.

The output led to a nationwide PPIE Implementation Program funding and facilitating public involvement activities in research launched in Autumn 2020.

PPIE IMPLEMENTATION PROGRAM

The PPIE Implementation Program (ppie.lbg.ac.at) aims to support 'active involvement' (Hayes, Buckland, and Tarpey 2012) of public members in research activities across different phases of the research cycle – from setting the agenda to interpreting data – and its governance. It supports public involvement activities with up to EUR 60.000 over a project period of 6-12 months implemented at Austrian research organizations and universities. Private and public Austrian research organizations were eligible for funding. Eligible costs included honorarium for public members, other direct costs, travel costs for public members and researchers, and 20% overhead costs. Personnel costs for researchers and research equipment were not eligible. Applicants were asked to submit a three-page application describing the societal impact, implementation plan, considered methods, and expected learnings of the PPIE activities. The first PPIE Pilot Call opened in September 2020 (total funding volume EUR 600.000) and accompanied support offers on the administrative and methodological implementation of the project idea prior to submission in form of individual consultations and webinars. A second PPIE pilot call was planned in September 2021 (total funding volume EUR 600.000).

In addition, and at the core of the program, we aim to build at LBG OIS Center an institutionalized support for PPIE projects located within LBG as well as at Austrian research institutions. This was supported with staff representing 1,5 full-time equivalents. This includes individual consultation and training opportunities in PPIE related topics and participatory methods, such as webinars and co-creation workshops with different stakeholder groups, as well as creating learning opportunities through a peer network. The peer network aimed to establish a PPIE community and embed public involvement in the Austrian research landscape and beyond. The support offers were available without charge for researchers and public members in Austria and assessed after each event with a questionnaire tailored to each support offer. Moreover, the PPIE Implementation Program – the funding instrument and support offers – will be externally evaluated.

As part of the Program evaluation, the evaluation of the funded PPIE projects and activities include views from all stakeholders that participated in the PPIE activities (researchers and members of the public). The projects' evaluation questionnaire addresses the following dimensions: quality of involvement, learnings from activities, future and sustainability of activities, scientific and societal impact of activities on individual and organizational level, implementation of activities, and satisfaction with the PPIE activities. The project evaluation is conducted once after the end of the project period.

TRANSDISCIPLINARY EXPERT EVALUATION PANEL

Based on previous experience of involving experts by experience (e.g., patients and citizens) in project steering and governance – for example, people with lived experience in a field of mental health (Kaisler & Paul 2019) – we established an independent and international expert panel including members of the public for assessing the PPIE funding applications. The expert panel aimed to assess and select the high-quality applications. It consisted of two scientific experts in the field of public involvement, a patient in the field of health, and two students with basic scientific background (16-30 years). We established characteristics for each group of experts (scientists, patient, students), which consisted of mandatory skills (e.g., fluent English for all experts, or lived experience in case of the patient) and desirable skills (e.g., experience in committee work for all non-scientific experts). To align the experts to the goal of the call, we organized a briefing meeting introducing the scope of the call, the assessment criteria and gave the experts the opportunity to get to know each other and to explain their respective relation to participatory science.

The transdisciplinary expert panel assessed the project proposals individually based on four criteria: quality of involvement, societal impact, implementation plan, and feasibility within the given time frame. The assessment focused on the participatory approach and its quality of involvement (Hayes, Buckland, and Tarpey 2012) rather than the scientific approach. After individual online assessments, the expert panel discussed outstanding PPIE project proposals in a selection meeting and recommended the highest rated projects for funding to the LBG Management Board.

In the selection meeting, we were able to observe good and respectful cooperation in which the experts were responsive to each other's perspectives and concerns. These different perspectives led to discussions

about 'high-quality' involvement activities and a consensus among the participating experts.

RESULTS OF THE FIRST PPIE PILOT CALL 2020

A total of 29 applications (62% female and 38% male applicants) were submitted from 15 different research institutions in Austria. Thereof four applications were not eligible for funding. More than half of the applications were submitted from other federal states than Vienna. In line with the scope of the call, most of the applications were thematically related to health sciences. The PPIE Pilot Call 2020 primarily addressed scientists in early career stages (24% PhD students and 41% PostDocs) from universities (34%), research organizations (28%), university of applied sciences (14%), private universities (7%), and public agencies (7%).

The expert panel recommended 11 of 25 eligible proposals for funding (81% female, 19% male) with a total amount of EUR 505.193. The success rate of female applicants increased (81%) compared to the application stage (62%). The successful projects addressed the following areas: medicine (46%), social sciences (45%) and psychology (9%). 46% PostDoc researchers, 27% PhD students, 18% professors, and 9% research administrators were granted. Their project ideas describe many different participatory approaches, such as co-creative workshops with stakeholders and members of the public (37%), the establishment of project steering and advisory boards including patients (27%), the co-development of questionnaires and research activities (27%), and involving patients as co-researchers in the research team (9%).

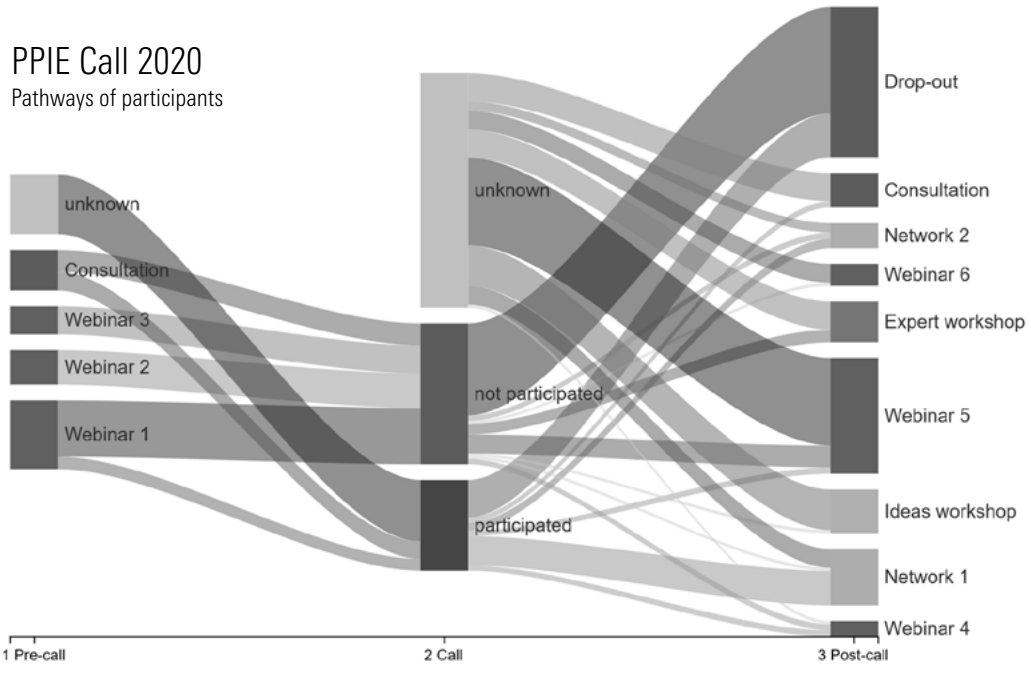
The expert panel decided not to use the entire funding volume of the call because some proposals did not convey the desired quality of involvement. More than half of the projects (55%) used support and consultation prior to the submission, thereof 60% succeeded in funding. In comparison, 45% did not take advantage of consultation, but still succeeded in funding. Preliminary results from the evaluation of the support offers indicate that the project leaders were satisfied with both, the funding instrument and support offers, and they expressed their interest in learning from the other peers.

EVALUATION OF SUPPORT OFFERS

The PPIE support offers aimed at building a second pillar alongside the funding instrument. The support offers provided low-threshold opportunities to get familiar with the topic and consult project ideas with the experienced researchers working with participatory methods. It is therefore primarily aimed at researchers, although other stakeholders were also welcomed to participate and take part in some of the offered activities (e.g., idea workshop with stakeholders). For this purpose, the PPIE Implementation Program offered a series of different activities to support the introduction and implementation of PPIE to researchers and to inform about important aspects or existing solutions regarding PPIE. Here, we analyze the eleven activities (including the PPIE Pilot Call 2020) that were conducted from September 2020 until June 2021. These activities vary from face-to-face consultation (n=18), webinars (n=6), idea workshop (n=1), PPIE Pilot Call 2020 (n=1), expert workshop (n=1) and network meetings (n=2). Due to the Covid-19 pandemic all activities were held online.

In total, we documented 248 interactions since the start of the program within this timeframe. An interaction is defined as: a person showed intention to participate in one of the support offers, such as by registering or by making an appointment. Therefore, a person may have several interactions. In 197 cases (79%) the interaction resulted in the participation of a person in an activity of the support offers.

The webinars accounted for the most interactions as the webinars were most frequently offered and had the lowest barrier to participate. On average, we had 23 interactions per webinar compared to around 15 in the other formats (consultation and pilot call excluded). Nevertheless, the attendance rate of the webinars was the lowest (66%, see Table 1) compared to the other activities.



Caption Figure 1. Pathways of participants' attending different support offers.

Figure 1 shows the initial and following interactions of participants in different support offers. It indicates that the webinars raised awareness about other PPIE support offers but had limited influence on submissions to the PPIE Pilot Call 2020. While about one third of the first webinar's participants also attend later activities, only a fraction of them applied for the PPIE call. The two content-related webinars (2 and 3) did not result in any further applications in the PPIE Call. This indicates that the webinar format did not attract potential applicants to the call. However, it generated broader attention for the support offers as well as further interactions in other activities. About half of the consultations (average duration about 60 minutes) prior to the call resulted in an application. It seems that this format – providing individual feedback to projects ideas – was more effective in terms of attracting researchers to apply. Further, more than half of the applicants also attended in a later activity of the support offers. This indicates that the funding instrument generated commitment to the support offers in the first round of the PPIE Pilot Call. A large proportion (Fig. 1 'unknown' column) of the participation in the call and other support activities did not result from prior interactions with the PPIE Implementation Program. These unknown participants may have been recruited via social media, newsletters, and information on topic-related platforms as well as word-of-mouth dissemination of information in the field. This is especially valid for capability building in the PPIE community in Austria. Based on the distribution of disciplines and topics submitted, we conclude that we have reached different participatory approaches in health research.

In total the 248 interactions led to a reach of 163 individual persons. On an individual level, 31 registered individuals (19%) never participated in any of the support offers, while 87 individuals (53%) participated in one and 45 individuals (28%) in two or more offered activities. Of these 45 individuals, 20 were identified as regular users who participated three or more times in the support offers. It remains open why 19% of users did not attend the activities.

The analysis shows that the group of no-shows was mainly interested in the webinar format. 42% of the no-shows worked at foreign research institutions and were consequently not eligible for the PPIE Call. Among the group of regular participants only 5% (n=1) were associated with a foreign research institution. The group of regular participants also made particular use of the PPIE consultation or tended to take part in more advanced formats such as the network meetings and the expert workshops. This highlights the importance of the funding instrument to encourage the rather fragmented community of different disciplines and participatory approaches to interact on a regular basis.

The webinars were designed as a low-threshold format and consequently generated the least commitment. Nevertheless, the webinars enabled the highest mobilization among people, while activities with a higher threshold (expert workshop, network meeting) required more incentives to participate. In the case of the PPIE Implementation Program, the prospect of funding may have played a role. In general, the different support offers led to a more diverse audience with different needs being addressed by the program.

Table 1 Evaluation of participants' satisfaction with different support offers.

	Webinar	Network	Consultation	Expert WS	Idea WS	Total
Total number of activities	6	2	18	1	1	28
Number of participants	91 (137)	26 (30)	24 (24)	13 (15)	14 (14)	168 (220)
Attendance rate	66,42%	86,67%	100,00%	86,67%	100,00%	76,36%
Number of survey respondents	17	7	7	11	10	52
Overall satisfaction event (1-5)	4,68	5,00	5,00	4,91	4,60	4,70
Satisfaction with learnings (1-5)	4,47	4,25	4,71	4,66	3,97	4,38

Note: Numbers in brackets indicate total number of registered participants for the activity. We used a 5-point Likert scale to assess the satisfaction of participants (1 not satisfied to 5 very satisfied). WS = workshop.

Table 1 shows the different activities of the support offers describing the attendance rate and the satisfaction of the participants. To adapt our support offers to the needs of the community and to improve the implementation, a small questionnaire ("participation check") was sent to participants after each activity. The participation check surveyed how comfortable the participants felt with the event, how well it was implemented and whether they were satisfied with the takeaways from the event. As the activities were conducted online due to the Covid-19 pandemic, the participation check was also conducted online, even though this may have had a negative impact on the response rate. For this reason, we have significantly shortened the questionnaire for the webinars, which has improved the response rates to some degree, although we lost some interesting information.

These findings indicate that the overall satisfaction of the participants across all activities was high, ranging between the scores four and five on a scale from 1-5 ('not at all satisfied' to 'very satisfied'). While satisfaction with the webinars was somewhat lower, the consultations and the peer network seem to be particularly well received. The open field in the questionnaires highlighted that the respondents appreciated the offer and their relevance. Respondents positively underlined that the webinars gave a short overview of the areas and that the topics were of interest. However, the respondents mentioned that less time was dedicated to networking with other participants which was frequently emphasized as impediment.

To bridge this gap, we offered two interactive settings in addition to the network meetings, i.e. the expert workshops and the idea work-

shops. The overall satisfaction in these two formats was high. The responds in the open fields emphasized that they were well suited for joint learning and cooperation opportunities as well as to gain knowledge from different perspectives.

Despite the very positive feedback in the open fields, the ideas workshop scored lowest in both overall satisfaction and takeaways. This shows the weakness of the survey design lacking information about motives and backgrounds of the participants. In case of the ideas workshop, the majority of participants were patients or citizens, which might evaluate the questions more critically than scientists who are more familiar with such surveys.

The consultation received the highest ratings in terms of both satisfaction and takeaways. Both in the survey and as feedback during the consultation, respondents emphasized the unique position of this offer in the field and its relevance for developing a participatory approach in research projects. This offer was less supportive of the networking of individuals in the field but was extremely effective in providing knowledge about the call and participatory research approaches. For the PPIE team, it was also a direct opportunity to learn about barriers and opportunities in the field and, if necessary, to follow-up with a tailored support offer (e.g., topic for expert workshop).

LEARNINGS AND CONCLUSION

With the PPIE Implementation Program, we aimed to introduce patient and public involvement in the Austrian research landscape as well as a new research-funding instrument by involving members of the public in the development of the activities, assessment of the project proposals and the overall evaluation of the program.

In the consultation sessions we experienced difficulties from applicants to describe their participatory approach. This often resulted in describing the scientific approach instead and left the expert panel with open questions in their assessment of applications. For this reason, we are adapting the application documents in the second call and expand the consulting activities. The latter is also important as non-research bodies, e.g., patient organizations and non-governmental organizations, are eligible for funding in the second call, which was recommended by the expert panel. Non-research institutions may need more support in preparing their applications, as they are usually less familiar with applying for funding. Therefore, we will make the application more accessible for public members by allowing applications in German in justified cases and introducing a video format additionally to the text-based application.

In the selection meeting of the transdisciplinary panel, we found that – despite the briefing – different understandings of active participation were negotiated. However, this was not a problem due to the open and collaborative atmosphere and mutual learning opportunity for experts. This shows the importance of creating an atmosphere where all participants can get involved, provide facilitation, and allocate enough time for discussions. Experts reported a high workload assessing all 25 applications in detail. We underestimated the effort required for less trained experts during the assessment. In the second call, we will allocate a maximum of ten applications for each expert for individual assessment.

Regarding the evaluation of the support offer, we learned that the questionnaire was not suitable for drawing conclusions on the PPIE Implementation Program and target group. At first, we decided to exclude personal data such as gender or institutional background etc. ensuring anonymity to the respondents and keeping the survey as short as possible. Also, we did not perceive these data as relevant for the evaluation of the activities. While this provides general feedback on the activity, it reduced our possibilities to evaluate the single activity connected to the PPIE Implementation Program as we lack knowledge about the respondents' perspective and the reasons why they participated. Therefore, in the second PPIE Pilot Call 2021, we revised the 'Participation Check' based on the experience gained so far including demographic data about participants, and feedback on the format. Nevertheless, there are some lessons we can draw from the participation check so far – especially in combination with the responses in the open fields. The evaluation of the support offers indicated that a pure focus on dissemination events (like webinars) is not sufficient to foster capacity building in community. It seems that there is a need for formats in which people work together and thereby come into direct contact, such as co-creative settings to jointly find solutions for the problems of others.

The different support offers were identified as strength of the PPIE Implementation Program, well perceived in the PPIE community and positively evaluated by the participants. However, the cross-linking between the different activities should be improved to increase the share of repeated participations and thus improve networking effects.

Despite the need for small adaptions in the second PPIE Pilot Call 2021, it seems that the funding instrument was appropriate and reflects a low threshold offering for researchers introducing public and patient involvement activities in their work. The PPIE Implementation Program is a first step towards establishing high quality public and patient involvement in research and an institutionalized PPIE support structure.

REFERENCES

Beck, S., Bergenholtz, C., Bogers, M., Brasseur, T.-M., Conradsen, M. L., Di Marco, D., Distel, A. P. et al. (2020). "The Open Innovation in Science research field: a collaborative conceptualisation approach." Industry and Innovation: 1-50. https://doi.org/10.1080/13662716.2020.1792274.

Boylan, A. M., Locock, L., Thomson, R., & Staniszewska, S. (2019). "About sixty per cent I want to do it": Health researchers' attitudes to, and experiences of, patient and public involvement (PPI)-A qualitative interview study. Health expectations : an international journal of public participation in health care and health policy, 22(4), 721–730. https://doi.org/10.1111/hex.12883

Bundesministerium für Verkehr, Innovation und Technologie und Bundesministerium für Wissenschaft, Forschung und Wirtschaft (2016). Open Innovation Strategie für Österreich. Wien: Bundesministerium für Verkehr, Innovation und Technologie und Bundesministerium für Wissenschaft, Forschung und Wirtschaft.

Bundesministerium für Wissenschaft, Forschung und Wirtschaft (2015). Wissenschaft und Gesellschaft im Dialog „Responsible Science". Wien: Bundesministerium für Wissenschaft, Forschung und Wirtschaft.

European Commission (2014). White Paper on Citizen Science. Brussels: European Commission, Brussels.

Hayes H, Buckland S, Tarpey M. (2012). Briefing notes for researchers: public involvement in NHS, public health and social care research. Eastleigh, England: INVOLVE 2012

Hecker, S., Wicke, N., Haklay, M., & Bonn, A. (2019). How does policy conceptualise citizen science? A qualitative content analysis of international policy documents. Citizen Science: Theory and Practice, 4(1).

Greenhalgh T, Hinton L, Finlay T, et al. (2019). Frameworks for supporting patient and public involvement in research: Systematic review and co-design pilot. Health Expect 2019;22:785-801. https://doi.org/10.1111/hex.12888 (accessed 3 May 2021).

Kaisler, R. E., Kulnik, S.T., Klager, E., Kletecka-Pulker, M., Schaden, E., Stainer-Hochgatterer, A. (2021). "Introducing patient and public involvement practices to healthcare research in Austria: strategies to promote change at multiple levels." BMJ Open 11 (8): e045618. https://doi.org/10.1136/bmjopen-2020-045618.

Kaisler R. E., Missbach B. (2020). Patient and public involvement and engagement in research – a 'how to' guide for researchers. Zenodo 2020. http://doi.org/10.5281/zenodo.3515811 (accessed 7 October 2021).

Kaisler R. E., Missbach B. (2020). Co-creating a patient and public involvement and engagement 'how to' guide for researchers. Res Involv Engagem 2020;6:32. https://doi.org/10.1186/s40900-020-00208-3 (accessed 7 October 2021).

Kaisler, R.E. & Paul, J.L. (2019). Evidence-based practice and policies for impact on mental healht of children and adolescents. Fteval Journal, 48, 116-121. https://doi.org/10.22163/fteval.2019.377

Nathan, S., Harris, E., Kemp, L., & Harris-Roxas, B. (2006). Health service staff attitudes to community representatives on committees. Journal of health organization and management, 20(6), 551–559. https://doi.org/10.1108/14777260610702299

National Institute of Health Research. Involve patients [website]. https://www.nihr.ac.uk/health-and-care-professionals/engagement-and-participation-in-research/involve-patients.htm (7 October 2021).

Ocloo, J., Garfield, S., Franklin, B. D., & Dawson, S. Exploring the theory, barriers and enablers for patient and public involvement across health, social care and patient safety: a systematic review of reviews. Health Res Policy Sys 19, 8 (2021). https://doi.org/10.1186/s12961-020-00644-3

Thompson, J., Barber, R., Ward, P. R., Boote, J. D., Cooper, C. L., Armitage, C. J., & Jones, G. (2009). Health researchers' attitudes towards public involvement in health research. Health expectations : an international journal of public participation in health care and health policy, 12(2), 209–220. https://doi.org/10.1111/j.1369-7625.2009.00532.x

AUTHORS

RAPHAELA ELISABETH KAISLER
Ludwig Boltzmann Gesellschaft Open Innovation in Science Center
Nussdorferstraße 64/2, 1090 Vienna, Austria
E: Raphaela.kaisler@lbg.ac.at
ORCID: 0000-0002-1891-3508

THOMAS PALFINGER
Ludwig Boltzmann Gesellschaft Open Innovation in Science Center
E: Thomas.palfinger@lbg.ac.at
ORCID: 0000-0001-6933-5114

KEYWORDS: Patient and public involvement, funding, support, participatory evaluation approach

SUPPORTING RESEARCHERS UNDER THREAT IN TODAY'S ACADEMIA

LESSONS LEARNT FROM THE EVALUATION OF THE PHILIPP SCHWARTZ INITIATIVE

TOBIAS DUDENBOSTEL

DOI: 10.22163/fteval.2022.552

SHORT SUMMERY

The Philipp Schwartz Initiative (PSI) is a relatively new program of the Alexander von Humboldt Foundation (AvH) that was launched in 2016 in close cooperation with the Federal Foreign Office. PSI enables universities and other research institutions in Germany to host foreign scientists who are exiled, displaced, and threatened by war and persecution in their own countries. As Philipp Schwartz fellows they are entitled to continue their research for a period of two years.

Technopolis Austria was tasked with an evaluation of the first four selection rounds of PSI. The evaluation aimed at taking stock of program implementation, collecting interim results, assessing goal attainment and to provide recommendations to further improve the program. The program aimed at developing structures within organisations hosting threatened researchers, at integrating fellows into research to increase career perspectives, as well as at raising awareness and at sharing information and facilitate networking within German Academia. To our knowledge, this was the first evaluation of a comparable initiative.

Our contribution answers the following questions: First, how to best cater for the specific features of the program and the program beneficiaries in the design of the evaluation methodology? Second, what kind of methodological challenges did we encounter and what mitigation strategies were implemented? Third, what were success factors that enabled the program to reach its goals and what were barriers? And fourth, on a more general level, how is a program like PSI positioned within the AvH, German Academia and how can it contribute to safeguarding academic freedom?

To answer the evaluation questions, an evaluation concept combining qualitative and quantitative elements was developed and discussed with the AvH. There were several specific methodological challenges to overcome (building trust in the field of beneficiaries, privacy, data protection). The evaluation shows that the program objectives have been achieved to a large degree. We identified several success factors of the program such as program design, designated project structure, quick and flexible program administration, following a sensible division of tasks among stakeholders and lastly, community building and engagement. Barriers identified were personal difficulties (e.g., migration, threat, administration of e.g., refugee status, family, language, psychological distress). On a project level, mentors had to invest a high degree of personal time and commitment with comparable less support. On institutional level, psychological support was difficult to offer. Most important areas for improvement were defining better the role of mentors and increasing their support and/or exchange, funding the fellows through contracts (not stipends), and minor changes in monitoring.

PSI positions itself in an environment of on the one hand longstanding initiatives by NGOs and, on the other hand, various programs launched recently by research funding organizations such as the AvH. Implementing the program changed the AvH as well and provided an impulse to reflect on the German academic system, but PSI is also a contribution towards providing R&I resilience since it allows fellows to continue their research.

ACADEMIC FREEDOM AND THE PHILIPP SCHWARTZ INITIATIVE

With recent geopolitical developments, (re)emerging crises and a surge of authoritarian tendencies even in democratic countries such as the USA, Hungary or Poland the topic of academic freedom has been brought on the agenda. In a recently published study, Kinzelbach et al (2021) show how Academic Freedom is decreasing in countries such as Poland or even the USA, while being already critically low in countries such as China or Turkey.

This is, however, not the first time that relations between state, researchers and academia are changing: During the Age of Enlightenment, scholars sought to delimit themselves from state and church and claimed "libertas philosophandi", or the right to philosophize (Hoye, W.J. 2009). In the middle of the 19th century, liberal students demanded freedom of teaching and learning, resulting e.g. in the Austrian-Hungarian Staatsgrundgesetz of 1867 postulating the freedom of science and its teaching: "Die Wissenschaft und ihre Lehre ist frei." And lastly, in the years of the National Socialist regime, academic freedom was severely limited with, among others, researchers being expelled or even murdered, fields of research restricted and the institutional autonomy of universities reduced.

Therefore, Academic Freedom is typically defined comprising an individual and an institutional dimension. Vrielink et al. (2011) formulate:

- "Far-reaching individual rights to expressive freedoms for members of the academic community (both staff and students)

mainly as free enquirers, including the freedom to study, the freedom to teach, the freedom of research and information, the freedom of expression and publication (including the 'right to err'), and the right to undertake professional activities outside of academic employment;

- Collective or institutional autonomy for the academy in general and/or subsections thereof (faculties, research units, etc.). Said autonomy implies that departments, faculties and universities as a whole have the right (and obligation) to preserve and promote the principles of academic freedom in the conduct of their internal and external affairs."

With the Philipp Schwartz Initiative (PSI), the Alexander von Humboldt Foundation (AvH) combines individual and institutional aspects of Academic freedom. PSI is a relatively new program that was launched in 2016 in close cooperation with the German Federal Foreign Office.[1] PSI enables universities and other research institutions in Germany to host foreign scientists who are exiled, displaced, and threatened by war and persecution in their own countries. As Philipp Schwartz fellows they are entitled to continue their research for a period of two years. The period can be prolonged once for another year, if necessary.

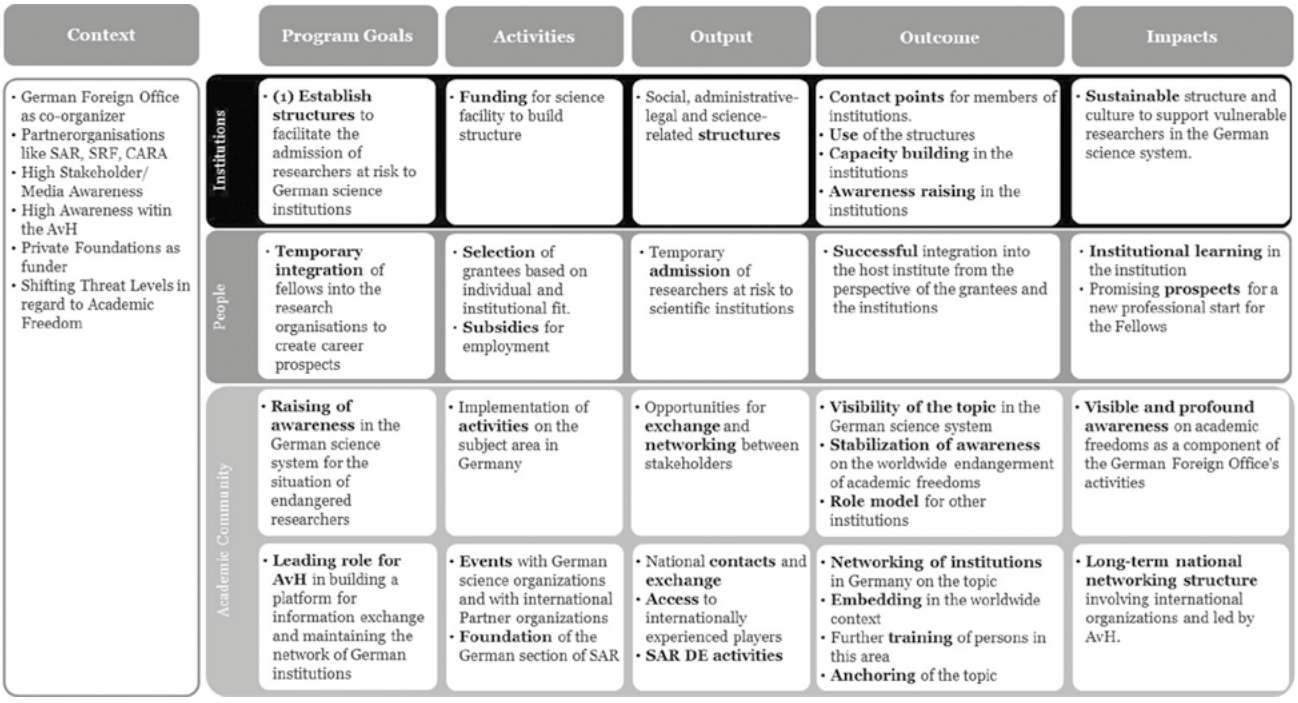
Interested host institutions can apply and present the AvH with suitable fellows as well as a Mentor who supports the scientific integration at the institution. They also need to present proof that the fellow is under threat, e.g. through an assessment carried out by a specialised organisation.[2]

Hosting institutions receive funding to award Philipp Schwartz fellowships and to establish structures that support the integration of the threatened researchers into the host institution. Moreover, the AvH fosters awareness raising activities and networking through accompanying events.

Working with threatened researchers is new for the AvH as well as for most German science organisations. In contrast to other AvH programs, PSI does not focus on scientific excellence, but on supporting foreign researchers under threat.

To our knowledge, this was the first evaluation of such an initiative. The program aimed at developing structures within organisations hosting threatened researchers, integrating fellows into research to increase career perspectives, as well as raising awareness and sharing information and facilitating networking within German Academia (see figure 1):

Figure 1: Logic Chart of PSI
Source: AvH, Technopolis



1 https://www.humboldt-foundation.de/bewerben/foerderprogram/philipp-schwartz-initiative
2 E.g. Scholars at Risk (SAR) or the Council for At-Risk Academics (CARA).

Technopolis Austria was tasked with an evaluation (Dudenbostel & Warta 2020)[3] of the first four selection rounds of PSI. The evaluation took stock of program implementation, collected interim results, assessed goal attainment and provided recommendations to further improve the program. In addition, the evaluation process also included an analysis of the fellows in terms of socio-demographic, legal and social characteristics, gather initial experiences from the beneficiaries and provide a systematic comparison of PSI to other relevant funding initiatives and programs.

This contribution, based on experiences made during the project, focusses on what is important when evaluating a programme for researchers under threat and thus, on the following questions:

- First, how to best cater for the specific features of the program and the program beneficiaries in the design of the evaluation methodology?
- Second, what kind of methodological challenges did we encounter and what mitigation strategies were implemented?
- Third, what were specific success factors that enabled the program to reach its goals and what were barriers?
- And fourth, on a more general level, how is a program like PSI positioned within the AvH, German Academia and within the discourse on academic freedom?

METHODOLOGY DESIGN AND CHALLENGES

The evaluation was guided by the question "what works, for whom, and under which circumstances?" The "Who's" that interested us were the researchers under threat on the one hand and the hosting institutions on the other hand. Within the first four selection round, PSI supported 162 fellows of which about 58% were from Turkey and another 30% from Syria. These were hosted by 48 German universities, 16 research institutes and nine Universities of Applied Sciences.

Due to the fact that this was the first evaluation and also one of the first studies in this specific context of researchers under threat, we needed to ensure that we developed a sound understanding of the researchers under threat and their specific contexts and backgrounds early on in the project. To that end, we strengthened the qualitative approach especially in the inception phase of the evaluation by organising focus groups with PSI fellows. Additionally, the evaluation team participated in the "Forum for Academic Freedom", an event organized by the Alliance of Science Organizations under the leadership of the AvH, to increase their knowledge of host institutions and important stakeholders as well as on the challenges, matters and impressions they already shared on the event. Alongside the focus groups, we conducted interviews with stakeholders, fellows, and mentors as well as with representatives of host institutions, some of which were developed into small case studies highlighting various aspects of program support. Interviews with fellows, mentors and representatives of host institutions were conducted face to face and on site. The collected qualitative evidence was further substantiated with a standardized only survey of the same target group. A media analysis of program communication as well as a comparison with other, national, and international programs concluded the methodological work.

A specificity of the program is the target group of researchers under threat. In praxis, PSI fellows had to leave their home country for several reasons, the most frequent one being that they were limited in their research due to their political views (n=66). Other, less frequent reasons were the destruction of infrastructure or equipment (n=29), their religion, ethnicity or sexual orientation (n=23) or their research topics (n=22).[4] However, for a number of fellows in particular from Turkey, political tensions were also present after arriving in Germany, for a few fellows even within the PSI community (see table 1):

Table 1: Political Tensions affecting PSI fellows

	Syria		Turkey		Other		Total	
	n	in %	n	in %	n	in %	n	in %
Yes…	14	40%	41	79%	4	31%	59	59%
...amongst people from my home country	11	31%	34	65%	3	23%	48	48%
...amongst PSI fellows		0%	4	8%		0%	7	7%
...the tensions are independent from place of origin	3	9%	3	6%	1	8%	4	4%
No	16	46%	5	10%	7	54%	28	28%
Not applicable	5	14%	6	12%	2	15%	13	13%
Total	35	100%	52	100%	13	100%	100	100%

Source: Survey amongst PSI fellows (n=100) Question: In your experience, are the political tensions that you know from your home country (if applicable) tangible in Germany as well? Single choice.

3 Dudenbostel, T.; Warta, K.: Evaluation der Philipp Schwartz-Initiative der Alexander von Humboldt-Stiftung. Endbericht (2020). Siehe: https://www.humboldt-foundation.de/fileadmin/Entdecken/Zahlen_und_Statistiken/Evaluation_der_Philipp_Schwartz-Initiative/ergebnisse_evaluation_psi_lang.pdf

4 Survey amongst PSI Fellows (n=101). Multiple answers possible.

Due to the nature of working with fellows under threat, i.e., people who find themselves in personal danger, specific challenges emerged. First, there was the question whether fellows would participate in our research at all. Second, if they would participate, how open could the evaluation team expect them to be? And third, when working with personal data, how to secure the data shared with us and how to make sure that information reported can be anonymized effectively?

Regarding participation in our research, feedback to our focus groups suggested that indeed, some fellows did not participate at all or did not feel safe sharing experiences or opinions in the group. Nevertheless, from the perspective of the evaluation team, the discussions that took place were informative and useful. To increase the participation of fellows in further research steps, in the end, trust was the most important factor. To increase the fellows' trust in our work, once more the qualitative and thus personal contact with them during the focus groups and field visits turned out to be crucial; jointly with the high reputation of the AvH among fellows. Additionally, mentors and institutions were important intermediaries for us. Apart from the feedback received to the focus groups, interviews and the anonymized online survey yielded the expected results.

While anonymization of data and data protection are important and guaranteed in all our projects, seldom are the stakes so high. That meant that for this project, data protection methods were further intensified by opting to priority data security over data protection: e.g., by minimizing the number of copies of the same data stored, by extremely limiting (internal) data availability, storing sensitive data only locally and by adding password protection on several layers. In terms of anonymization of case vignettes, which had an illustrative and explanatory function, we opted to proceed in the following way: first, fellows were asked for their consent to participate, the case vignettes focused only on very specific parts of their experience and its relevant context (while not providing much other information on the fellows behind the stories) and lastly, the information that was provided was mixed up between the cases.

RESULTS: PROGRAM BARRIERS AND SUCCESS FACTORS

The evaluation shows that the objectives have been achieved to a large degree. We argue that the following aspects identified in our evaluation constituted success factors of the program:

First and most importantly on an individual level, the PSI support enabled researchers under threat to focus again on their research in a safe environment. For that, the program funding and support was essential. Second, by design, the program asks host institutions to name Mentors that are responsible for the scientific integration of fellows into the host institution (although in practice, Mentors often do much more). Against the backdrop of the various difficulties the fellows have encountered prior to the fellowship and within, the distribution of labor within the projects between fellow, mentor and host institution was important for the project success. Third, in the context of providing support to persons under threat, quick and flexible program administration is crucial, both in setting up projects and in administrating ongoing projects. Our evaluation showed that PSI performed much better in this regard then other research funding schemes. Fourth, the program was designed following a sensible division of tasks. Most importantly, it was a good choice of the AvH to have NGOs like the Council for At-Risk Academics (CARA) assess whether applicants are under threat and thus eligible for funding. CARA and other NGOs have decades of experience in this regard. A fifth success factor was the readiness of the AvH to engage in relevant communities and networks early on (like Scholars at Risk (SAR)) and in community building where networks were not yet existent (as e.g., in most of the German Academia at the time of the program launch).

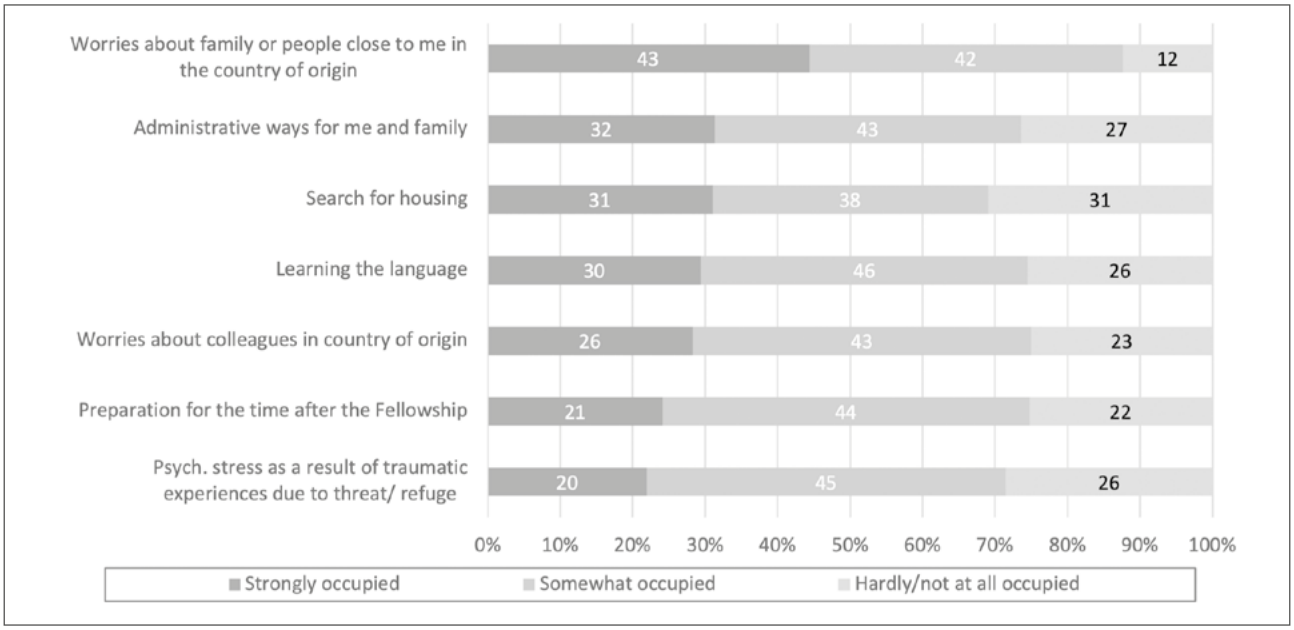
Nevertheless, a program supporting researchers under threat intervenes in a complex environment, and, in comparison to classic R&I funding schemes, in an environment with many more difficulties on a personal level. Barriers identified in our evaluation were: First, PSI intervenes in an incredibly difficult situation: researchers under threat are not mobile by choice but are forced to migrate – and that makes a difference! Often, they do not have advanced skills in the language of their new host country (72% overall, for researchers from Syria 42%), they often bring a family with at least one child (75%), in some cases the relocation itself was dangerous (about 25%) or at least arduous, and fellows and family reported being under psychological distress.[5]

In Germany, fellows are then introduced – often with little time to recover – to a highly competitive academic system where peers were supportive, but also competitors.[6] At the same time, fellows are occupied with organizing their stay administratively (e.g., apply for a refugee status), with finding residences, and organizing childcare or education for their families, etc. (see figure 2). Not all of these challenges are specific for researchers under threat when compared to what international researchers usually encounter – but for researchers under threat, they all come together at once.

5 Survey amongst PSI Fellows (n=101).

6 Fellows reported that in comparison to their home institutions, the scientific level at the host institutions was higher (more than 70%).

Figure 2: fellow's Occupation by selected Items



Source: Survey amongst PSI fellows (n=88) Question: How far were you occupied by other tasks or problems that distracted you from your research or that took time? Single Choice,

Second, as outlined above, a success factor on the project level are the Mentors of the fellows. However, mentors were often PIs or Professors and thus already very occupied with managing day to day teaching and research. They were mostly motivated to participate in PSI because they wanted to help a researcher at risk (more than 85% agreed strongly).[7] To fulfill their role in the projects, Mentors had to invest a high degree of personal time and commitment. Nevertheless, about 90% of the Mentors answered that they would consider being a Mentor again in the future. That is why, third, it was a challenge both for the mentors and for the fellows to be able to allow for sufficient self-care as well. And fourth, while being scarce in general, psychological support for refugees was often lacking or at least not well known on an institutional level. In fact, a high share of fellows reported that they need firstly more information events on German residence law (more than 40%) and secondly, offers of psychological support in situations of stress (about 38%).

Based on the evidence collected in our evaluation, we identified several areas for potential improvements. The most important are: First, the role and tasks of mentors should be defined better, accompanied by an increased exchange of experience among current and potential mentors. Since several comparable programs have mentors, the AvH should act as a platform for this activity. Second, the way the fellowships are funded should be better adapted to the needs of the fellows. In line with other programs of the AvH, PSI used stipends to channel funding to the individual fellows. While stipends provide flexibility as they only concern the foundation and the fellows directly, in many cases, fellows perceived them as hindering their integration at their host institution, as most of the other researchers had direct contracts with the host institution. Adding to the feeling of "not belonging as much to the host institution", stipends also added to the administrative burden of the fellows as they had to organize e.g. their social security independently. Third, the lump sum that supports the host institutions – intended to help host institutions to develop and/or provide support structures for researchers under threat overall – was identified as an important mechanism to facilitate integration and for community building at and beyond institutions. The way the lump sums are used should therefore be monitored. Lastly, the involvement of stakeholders from industry should be strengthened to help fellows increase career prospects outside of academia.

CONCLUSION AND OUTLOOK

During our evaluation, it quickly became obvious that PSI or similar programs were at the time novel for the AvH, Germany and continental Europe. However, several comparable programs were launched at the same time on an institutional or regional level in Germany, but also in France and, to some degree, within EU-programs. The positioning took place against the backdrop of the increased refugee occurrence in 2015, but many see themselves in a wider historical context of fostering academic freedom on an individual level. The two existing NGO support schemes stem from the 1930ies, when individual academic freedom was under threat on a larger scale in Germany and Austria and when standard works of sociology or philosophy of science were authored, e.g., Robert K. Merton's essay on The Normative Structure of Science (1942).

As the evaluation team participated in several events, it became also evident that there was a high level of (also personal) commitment within the emerging community of institutions and persons engaging with the

7 Survey amongst PSI Mentors (n=71).

topic. The program focused on helping foreign researchers under threat by enabling institutions to host them. Other objectives, especially the development of structures at the host institutions were also considered important, but less central to program beneficiaries and therefore, did not receive the same attention at all institutions. Furthermore – in an area of great difficulty and at a moment of great potential for distraction – PSI directs the fellows' attention to the time after the fellowship.

We argue, however, that implementing the program changed the AvH as well and provided an impulse to reflect on the German academic system. Engaging with researchers under threat means taking a different perspective on what public funding in an academic context should achieve: PSI is not about funding excellent science, but about supporting researchers under threat to conduct their research in safety and to provide a perspective towards career possibilities, possibly beyond academia. Therefore, PSI differs from other research funding schemes overall and in the AvH portfolio. In many discussions with policy makers, stakeholders, and researchers on the topic, it was also argued that this perspective is underdeveloped within German academia overall which is highly geared towards enabling excellent research through competition, but might neglect negative effects on the individual researchers competing.[8]

PSI can also be seen as a program safeguarding individual academic freedom on a global level, since many of the fellows fled their home countries because their research or research topics were seen as adversarial by those in power. Other researchers fled war or destruction. In both cases, PSI can allow for the continuation of research trajectories, although there is evidence that some researchers change their research topics, e.g. following their personal experiences with flight, migration, and oppression. Overall and in the long run, more than 80% of PSI fellows would like to continue their research and stay in academia.

This is one of the parallels of current fellows with the name giver of the program, Philipp Schwartz. Schwartz was a professor of pathology in Frankfurt and had to flee Germany in 1933 due to the Nazi terror. In Switzerland, he founded the "Notgemeinschaft deutscher Wissenschaftler im Ausland", a support organization for German refugee scientists. For many of those and including Philipp Schwartz himself, the organization found a new place of residence to work in safety, in a country that was, back then, rapidly modernizing its research and higher education system: The Republic of Turkey (Kreft 2015). There, Philipp Schwartz also shifted his research focus towards social medicinal topics. That was decades ago, though. Today, as indicated above, 60% of the researchers under threat supported by PSI until August 2018 stemmed from Turkey.

REFERENCES

Kreft, Geralt (2015): Neuroscientists rescuing refugee scholars: Three Founders of the Notgemeinschaft in Zurich, 1933, in Swiss Archives of Neurology and Psychiatry 2015; 166(8): 293–297, https://www.degruyter.com/downloadpdf/j/nf.2007.13.issue-1/nf-2007-0106/nf-2007-0106.pdf.

Vrielink, J., Lemmens, P., Parmentier, S. Academic Freedom as a Fundamental Right. Procedia - Social and Behavioral Sciences. 13 (2011). 117-141. 10.1016/j.sbspro.2011.03.009.

Hoye, J. William (2009). Wurzeln der Wissenschaftsfreiheit an der mittelalterlichen Universität, in: Wissenschaftsfreiheit, Reihe: Braunschweiger Beiträge zur Sozialethik Band 2, hrsg. von H.-G.Babke (Frankfurt am Main: Peter Lang 2009), 19–47

Kinzelbach, K., Saliba, I., Spannagel, J., Quinn, R.: Free Universities, Putting the Academic Freedom Index into Action (2021). https://www.gppi.net/media/KinzelbachEtAl_2021_Free_Universities_AFi-2020.pdf

AUTHOR

TOBIAS DUDENBOSTEL
Technopolis Austria
Rudolfsplatz 12/11, 1010 Wien
E: Tobias.dudenbostel@technopolis-group.com

KEYWORDS: Academic freedom, refugee scholars, AvH, privacy, HEI, PRO

8 See for example the current discussions on the German Wissenschaftszeitvertragsgesetz under the hashtag #IAmHanna.

CAN WE PREDICT SUCCESSFUL MARKET INTRODUCTION USING ON-GOING R&D EVALUATION DATA?

SHUMPEI MIYAJIMA, TOSHIYUKI ISSHIKI, MOTOSHI KUNUGI AND SHIN UESAKA
DOI: 10.22163/fteval.2022.553

ABSTRACT

This study aims to find reproducible correlations/causality between the evaluation data of ongoing R&D projects funded by NEDO and the ex-post monitoring data of actual commercialisation achievement by those projects. The understanding of the results of this study will be used for designing our R&I policies for the next era as a funding agency by, for example, promoting more effective schemes which will eventually increase our contribution to society.

The results showed positive correlations between the assessed grade for sections of the evaluation and the commercialisation status, indicating the possibility of identifying those projects that need management revision before the extended R&D activities by the companies.

INTRODUCTION

ABOUT NEDO

Following the two oil crises of the 1970s, New Energy and Industrial Technology Development Organization (NEDO) was established in 1980 to promote the development and introduction of new energy technologies. Since then, NEDO has become one of the largest public research and development management organisations in Japan, and it works with the government to implement economic and industrial policies.

In this capacity, NEDO undertakes technology development and demonstration activities to carry out the two basic missions of addressing energy and global environmental problems and enhancing industrial technology by integrating the combined efforts of industry, academia, and government.

THE EVALUATION SYSTEM IN NEDO

NEDO has established and been applying its evaluation system for two decades. Figure 1 shows the overall scheme of the present evaluation and survey scheme for a typical 5-year project. Starting from the project planning stage, we have a set of four evaluation opportunities-chances for each project plus an extended survey for selected projects.

(1) An *ex-ante assessment* is performed when it is still at the project's planning stage to see if the project is worth being carried out. The results of an ex-ante assessment are fed back for refining the project plan and requesting the final budget scheme.

(2) A *mid-term evaluation* is performed typically once for a project, and the results are directly reflected onto the management of the project for the rest of the period.

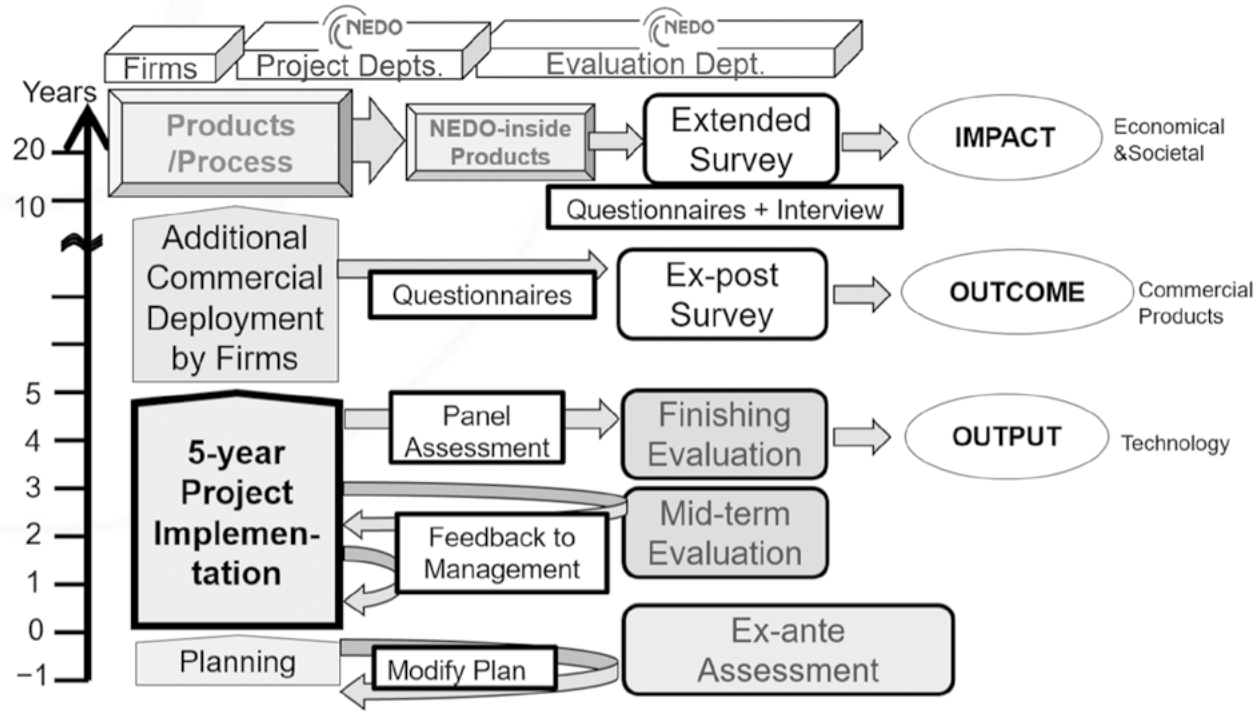
(3) A *finishing evaluation* is performed after the project is finished. The results of the finishing evaluation are often used as a reference for the planning of related new projects. An external panel of 5-10 evaluators is organised for each mid-term and finishing evaluation.

(4) After the end of the project, NEDO conducts ex-post surveys for up to six years (1, 3, 4 and 6 years after the end of the project). The NEDO evaluation department performs ex-post surveys, supervised by an external specific subcommittee, using questionnaires and interviews as the source data from the participant companies. The survey is necessary for the outcome evaluation, which assesses the post-project development by the participant companies and the resulting impact of the project on society.

(5) An *extended survey* is performed for selected projects that have produced products with economic or societal impact. We name these selected products as "NEDO inside products". As of 2020, 120 products are registered.

The outcomes and impacts of all NEDO projects are then used for accountability for taxpayers and for improving the project management system in general.

Figure 1. The overall scheme of the present NEDO evaluation and survey



OBJECTIVES OF THIS RESEARCH

As a funding agency for accelerating innovation, NEDO's responsibility is to maximise the outcomes of national projects through the commercialisation of their development results. To realise this mission, it would be beneficial if the activities after the completion of the funded project could be controlled by referring to the evaluated score of the project to facilitate commercialization.

The evaluation department of NEDO has set up a system to evaluate each project's output and assess the project's outcome after the completion of the project (as in Figure1).

This study hypothesises a correlation between the results of the finishing evaluation and the ex-post survey of each project. If so, it is possible to predict the expected extent of commercialisation from the evaluation results during or just after the project implementation. This research contains a new way of understanding data in that it analyses and uses the results of two systems of data, output evaluation and outcome assessment.

METHODOLOGY

We used all 334 NEDO projects completed between the years 2002 and 2013 with results of finishing evaluation as the population of the analyses, 178 of which also conducted outcome surveys for up to six years after completion.

There are two categories in the nature of NEDO projects: "standard" type and "basic" type. The standard ones aim to commercialise new products mainly through applied research and development during the funded project, while the basic ones are implemented starting from more fundamental research. Of 334 projects, 167 were standard type projects, and 167 were basic type ones. Of the 178 projects that completed both finishing evaluation and ex-post survey, 99 were of standard type, and 79 were of basic type.

(1) Finishing evaluation at the end of the project

Projects were evaluated at the end of the implementation (hence the name finishing evaluation) by a panel of five to ten evaluators selected from outside the organisation for each project. For each project, detailed evaluation items were set along each of the four viewpoints. NEDO's four evaluation viewpoints

Position & Significance,
Project Management,
R&D Achievement, and
Prospects for Practical Application.

According to the pre-defined criteria, each panel member marked between 0 and 3 for each viewpoint.

In addition to the scores, the evaluators provided detailed comments on each evaluation item. This paper does not treat comments, although an interesting textual analysis is expected.

(2) Ex-post survey

In NEDO's project system, several companies participate in one project to develop related product groups. A total of 684 firms participated in the 99 standard-type projects, and 441 companies participated in the 79 basic-type projects. Electronic surveys were conducted with these companies asking them about their R&D progress since the end of the NEDO project.

The survey questions include whether they were still developing the product, the current TRL (Technology Readiness Level) of the product development, and what factors they think might have contributed to the success or failure of product development. In addition, relevant questions related to the status of the product's TRL, such as the sales amount and the launching date, precede the TRL question itself to minimise the inaccuracy to some extent in the case the responsible staff for the product have changed since the previous survey.

The TRLs used by NEDO are by NEDO's definition. A simplified version used for this study is shown in Figure 2. In this study, stages 3 and 4 are combined to make a category "Practically Applied", and the percentage of products (companies) that reach the Practically Applied stage for each project is called the "commercialisation rate" of the project. In addition to these four TRL stages, some projects are discontinued at some point in the six years of the ex-post survey period, and these cases are counted as "discontinued". Also, using the answer to the first ex-post survey, the percentage of projects that did not immediately stop in-house development after the end of the funded project but at least continued then is called the "immediate continuation rate".

NEDO-TRL
NEDO TRL-1 Research: fundamental/elemental research
NEDO TRL-2 Technology development: research withtaken into consideration practical application/commercialization
NEDO TRL-3 Practical application: establishment of technologies for practical application/mass production
NEDO TRL-4 Commercialization: transactions in the market

Figure 2. NEDO TRLs. TRL-3 and TRL-4 combined define the "practically applied stage" in this study

RESULTS

First, trends were analysed within each of the two datasets, the finishing evaluation and the ex-post survey. Correlations between these two datasets were then examined.

(1) Results of the finishing evaluation

The distribution of finishing evaluation scores of all projects for the four viewpoints is shown in Figure 3 and Table 1. The mean score for the first viewpoint (Position & Significance) was the highest among the four viewpoints. This is because the position of the project is assessed at the end of the project when the project has already been running for five years, which increases the number of positive evaluations. It is a challenging issue whether to include this viewpoint in the finishing items.

The mean score for the fourth viewpoint (Prospects for Practical Application) was lower than for other viewpoints. One interpretation of this is as follows. During the implementation of a project, the management side tends to prioritise the achievement of direct development objectives, and relatively less consideration is given to the actual prospects after the end of the project.

Next, a comparison is made between the standard and basic types of projects for each evaluation viewpoint. T-tests showed a significant difference at the 5% level for the second viewpoint (Project Management), with the standard type having a higher score. As NEDO's R&D is aimed initially at applying already developed technologies, the effort to manage basic type projects may have been relatively weak.

Figure 3. Distribution of scores for four viewpoints

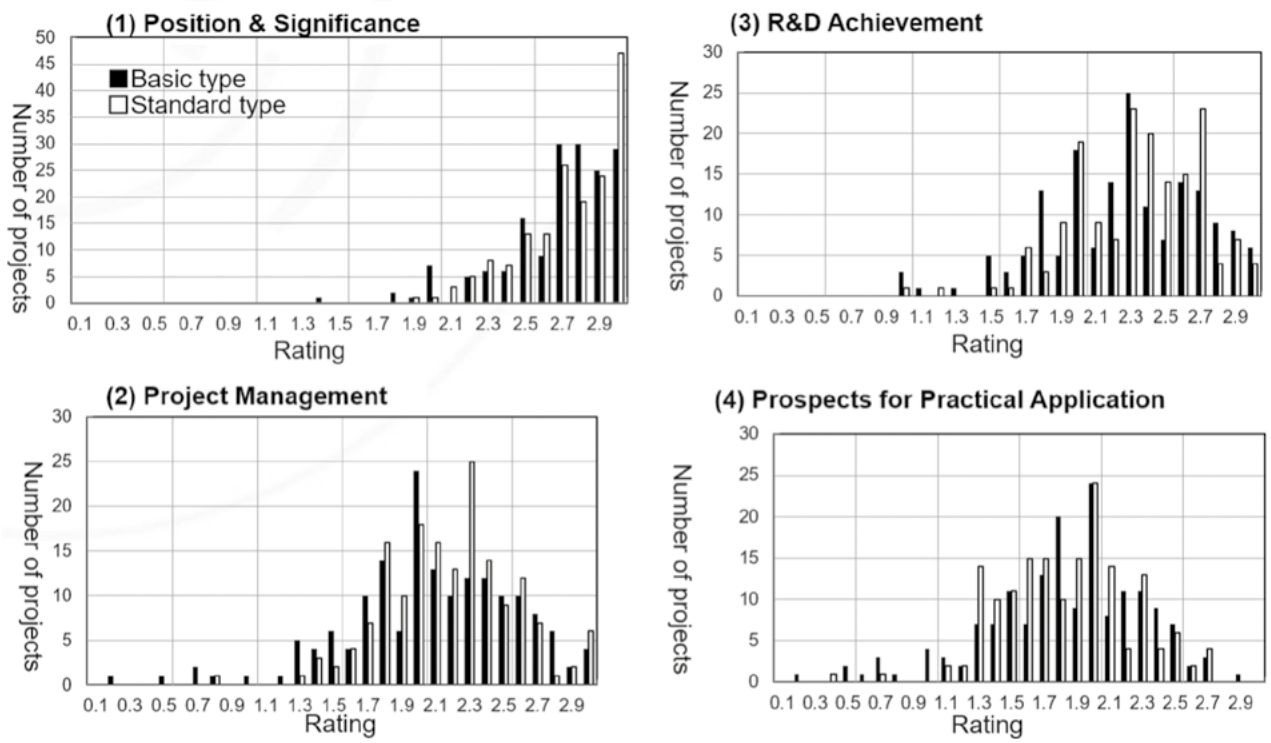


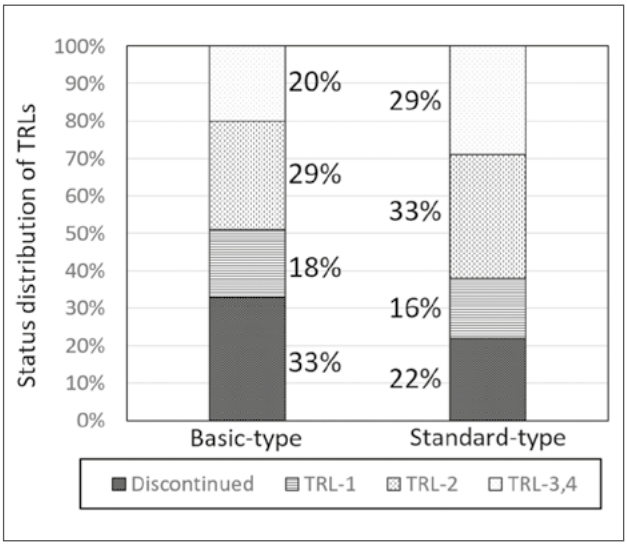
Table 1. Distribution of scores for four viewpoints (SD: standard deviation)

	(1) Position & Significance		(2) Project Management		(3) R&D Achievement		(4) Prospects for Practical Application	
	mean (median)	SD	mean (median)	SD	mean (median)	SD	mean (median)	SD
standard type projects (N=167)	2.73 (2.8)	0.26	2.18 (2.2)	0.38	2.34 (2.4)	0.36	1.84 (1.9)	0.40
basic type projects (N=167)	2.68 (2.8)	0.30	2.08 (2.1)	0.50	2.26 (2.3)	0.44	1.83 (1.9)	0.48

(2) Results of the ex-post survey

We examined the status of each product in the NEDO-TRL (1-4 and "discontinued") for the standard type and basic infrastructure type projects in the final surveys, which were done 6 years after the end of the project. The results are shown in Figure 4. The ratio of the commercialized (TRL 3 or 4) products to the total products is 29% for the standard type and 20% for the basic type. The percentage of discontinued products was 22% and 33%, respectively, with the basic model being the larger of the two.

Figure 4. Distribution of TRLs for standard and basic projects obtained from the ex-post survey six years after the end of the projects



The reasons for these results may be that since it generally takes longer to commercialise basic-type projects than standard-type ones, the percentage of abandonment is relatively high for basic-type ones due to lack of judgment at the end of the project. It makes sense to discontinue the project that cannot show its future.

(3) Correlation between the data sets of the finishing evaluation and the ex-post survey

The survey was conducted for 99 standard-type projects, for which both finishing evaluation and ex-post survey data were available.

The variables of the finishing evaluation selected for the correlation test were the scores given by the panel on each of the four viewpoints. For the ex-post survey, we used the number of companies in each project, the commercialisation rate, the immediate continuation rate at the end of the project and the abandonment rate after six years as the variables.

The correlation results are shown in Table 2. First, there is a positive correlation between the practical application viewpoint score of the finishing evaluation and the practical application achievement rate of the outcome survey with a 1% probability of significance. Next, there is a negative correlation between the score on every viewpoint of the finishing evaluation and the discontinuation rate of the ex-post survey at a 1% or 5% significance. Therefore, to some extent, it is possible to predict the likelihood of future commercialisation based on the finishing evaluation scores.

Table 2. Correlations between finishing evaluation and ex-post survey results of standard-type projects

(* and ** refer to 1% and 5% probability of significance, respectively)

		Finishing evaluation scores			
Ex-post survey scores		(1)Position & Significance	(2)Project Management	(3)R&D Achievement	(4)Prospects for Practical Application
	Number of companies	0.04	0.03	0.04	0.06
	Commercialisation rate	0.01	0.15	0.01	0.28**
	Continuation rate	0.14	0.00	0.15	-0.05
	Immediate discontinuation rate	-0.22*	-0.22*	-0.24*	-0.33**

DISCUSSION 1:

COMPLEMENTARITY BETWEEN FINISHING EVALUATION AND EX-POST SURVEY

There is a difference in the robustness of the approach for the results between evaluation, which assigns grades according to predetermined evaluation criteria, and ex-post survey, which is a set of self-reported answers to questionnaires. In evaluation, the project is assessed based on a causal relationship between the outputs and the implementation of the project rather than based on chance. Surveys based on self-reported data are generally considered insufficient to ensure a causal link between the implementation of the R&D project and its outcomes.

The correlations between the finishing evaluation and the ex-post survey shown in this study are not causal in themselves. However, suppose the finishing evaluation shows a causal relationship between project implementation and output expression. In that case, the correlation between extended R&D activities after the end of the funded project and outcomes found later in the ex-post survey is also assumed to have a causal element. It is unclear within the scope of this study how to quantitatively demonstrate that the correlation between finishing evaluations and outcome surveys has some causality, but if this hypothesis is correct, it would further enhance the usefulness of outcome surveys after the end of the programme. It would make the feedback efforts described below more meaningful.

DISCUSSION 2:

FEEDBACK AND REFLECTION ON CURRENT PROJECT MANAGEMENT

Based on the above concept, it is understandable that the outcome survey results could be reflected in the design of other similar projects. For example, an attempt could be made to increase the rate of future commercialisation by designing and managing projects in such a way as to raise the grade of the fourth viewpoint in the mid-term evaluation during project implementation.

NEDO has established a "Management Guideline" as a manual on project implementation for project managers. The Management Guideline focuses on some of the evaluation items in the finishing evaluation and recommends starting concrete efforts to tackle the items two years before the end of the project. For example, the company in the project should identify the department within the company that is responsible for commercialisation and set up a system for exchanging views with the project manager to clarify issues such as mass production technology and marketing. The idea is to increase the probability of successful commercialisation of the project by being aware of these items two years before the end of the project. This is evidence-based policymaking at a micro-level.

DISCUSSION 3:

USE IN ORGANISATIONAL EVALUATION

The results of finishing evaluations and ex-post surveys directly assess the project and the participant company but are not limited to these. The complete and accumulated results of the combined evaluations and surveys can serve as key performance indicators (KPIs) for the programmes and the organisation itself, which encompass the projects. It is also a reaffirmation of the robustness of the organisation's evaluation system.

Currently, NEDO has two indicators and corresponding targets for R&D performance among its organisational goals:

(1) The average commercialisation rate of projects that reach the fifth year after the completion between 2018 and 2022 should be at least 25%.
(2) 50% or more of the projects completed during the period mentioned above should achieve a score of 2.0 or higher on the 0-to-3-scale for "Prospects for Practical Application " at the finishing evaluation.

Although there is a five-year difference in the period covered, the above two indicators should be highly consistent, as this study has shown the robustness of the evaluation system. Let's consider the correlation between the distribution of scores on the Prospects for Practical Application viewpoint of the finishing evaluation and the distribution of achievement of practical application after six years in the ex-post survey. The two target values are expected to correspond roughly.

Looking at the past performance of these target values, out of 1,125 companies (standard type and basic type) in 178 projects for which finishing evaluation and ex-post surveys were completed, 284 (25.2%) achieved commercialisation. The mean value of the "commercialisation prospects" of the 334 projects for which a finishing evaluation was completed was 1.84. Both were close to the target value.

DISCUSSION 4:

MID-TERM EVALUATION

NEDO also conducts a mid-term evaluation during the implementation of long-term projects. However, the correlation between the mid-term evaluation and the outcome survey is not as straightforward as in the case of the finishing evaluation. One reason for this may be the distance between the evaluation and survey periods.

DISCUSSION 5:

FUTURE ISSUES

In addition to the 4-level scoring, detailed comments by the evaluators are collected in the finishing evaluation. The evaluators classify the comments according to the corresponding evaluation viewpoint and the positive/negative nature of the sentence. With this dataset of comments, statistical processing can be carried out to analyse the tendency of the comment on the above classification. In recent years, text mining analysis methods have been partly established in the Japanese language, and micro-analysis of the comments could be considered.

Our mid-term evaluation aims not only to assess projects, such as scoring and ranking, but also to adjust the project's orientation. Accordingly, NEDO has established a procedure to reflect evaluator remarks from the mid-term evaluation, and the relationship between the correction procedure and commercialisation is to be investigated.

Presently the survey also collects data on the amount of product sales as quantitative output data. However, a more appropriate quantitative criterion for the outcome is the value added by the project, to which sales figures are only a rough guide. To measure added values, it is necessary to establish a baseline before the project starts or apply equivalent counterfactual analyses.

The results of finishing evaluations and ex-post surveys such as those analysed in this study are likely to vary depending on the technical field of the project and the size of the companies involved. These will be analysed separately.

The details of the methodology in the finishing evaluation and the ex-post survey vary somewhat from year to year, particularly regarding the evaluation indicators. Because of the relatively significant changes in NEDO's evaluation policy in 2013, we have analysed projects for which finishing evaluations were carried out before 2013 in this study. It will be necessary to track later situation in the same way to examine changes over time.

CONCLUSIONS

We analysed NEDO's finishing evaluation results for output assessment and ex-post surveys for outcome assessment.

The finishing evaluation showed differences in the averaged evaluation score by the viewpoints of the evaluation and the characteristic of the project. Similarly, the ex-post survey revealed differences in the distribution of the TRL levels at the end of the survey by the characteristic of the project.

The analysis comparing the finishing evaluation results and ex-post survey results showed correlations between a particular viewpoint axis in the evaluation and indicators derived from the ex-post survey.

The results imply the possibility of utilising the result of the finishing evaluation for the management companies' R&D activity after the end of the project.

REFERENCES

S. Altuntas et al., "An evaluation index system for prediction of technology commercialization of investment projects", 2nd International Fuzzy Systems Symposium, 327 (2012)

J. Cho et al., "Development of a new technology product evaluation model for assessing commercialization opportunities using Delphi method and fuzzy AHP approach", Expert Systems with Applications, Volume 40, Issue 13, 5314 (2013)

M. Dziallas et al., "Innovation indicators throughout the innovation process: An extensive literature analysis", Technovation, Volumes 80-81, 3 (2019)

C. Karaveg et al., "Evaluation model for research and development commercialization capability", Production & Manufacturing Research, Volume 2, Issue 1 (2014)

J. Hyvärinen, "New evaluation framework in Finnish innovation policy", fteval Journal for Research and Technology Policy Evaluation, Issue 47, 41 (2019)

T. Isshiki et al., "Evidence-based planning for the medium-long-term R&D project management", Evaluation2018, American Evaluation Association, Cleveland (2018)

T. Isshiki et al., "Quantitative analysis on impact resulting from public R&D investment", European Forum for Studies of Policies for Research and Innovation, Paris (2018)

P-B. Joly et al., "Research impact assessment: from ex post to real-time assessment", fteval Journal for Research and Technology Policy Evaluation, Issue 47, 35 (2019)

M. Ueyama et al., "Impact evaluation of R&D support for SMEs and start-ups and its feedback on project management", fteval Journal for Research and Technology Policy Evaluation, Issue 47, 87 (2019)

V. Zemlickienė, et al., "Evaluation of the expediency of technology commercialization: a case of information technology and biotechnology", Technological and Economic Development of Economy, 26, 271 (2020)

AUTHORS

SHUMPEI MIYAJIMA
New Energy and Industrial Technology Development Organization (NEDO)
1310 Omiya-cho, Saiwai-ku, Kawasaki City,
Kanagawa 212-8554 JAPAN
E: miyajimaspi01@nedo.go.jp
ORCID: 0000-0002-3633-1608

TOSHIYUKI ISSHIKI
New Energy and Industrial Technology Development Organization (NEDO)
1310 Omiya-cho, Saiwai-ku, Kawasaki City,
Kanagawa 212-8554 JAPAN
E: isshikitsy@nedo.go.jp

MOTOSHI KUNUGI
New Energy and Industrial Technology Development Organization (NEDO)
1310 Omiya-cho, Saiwai-ku, Kawasaki City,
Kanagawa 212-8554 JAPAN
E: kunugimts@nedo.go.jp

SHIN UESAKA
New Energy and Industrial Technology Development Organization (NEDO)
1310 Omiya-cho, Saiwai-ku, Kawasaki City,
Kanagawa 212-8554 JAPAN
E: uesakasin@nedo.go.jp

KEYWORDS: Evaluation, survey, practical application, outcome, impact, correlation

RANDOMISED CONTROLLED TRIALS AND OTHER EXPERIMENTAL APPROACHES IN THE AUSTRIAN RESEARCH PROMOTION AGENCY (FFG) – EXPERIENCE, LEARNINGS AND OUTLOOK

TESS LANDON AND HARALD HOCHREITER

DOI: 10.22163/fteval.2022.554

1 ABSTRACT

The objective of this paper/presentation is to highlight how experimental approaches, specifically Randomised Controlled Trials (RCTs), can be leveraged to evaluate and measure the impact of new programmes, support programme development and test new services in funding and innovation agencies. RCTs are seen in many facets of public policy, however RCTs as a method for innovation agencies to evaluate new initiatives is relatively new.

We present three RCTs implemented in the Austrian Research Promotion Agency (FFG) that have received funding from the European Union's Horizon 2020 research and innovation programme. The trials are implemented to evaluate the effectiveness of new measures intended to help strengthen R&I in start-ups and SMEs. Through these three examples, we aim to demonstrate the advantages in which RCTs can augment the evaluation of new services as well as challenges that come with implementing RCTs. For one RCT, we will present final results. Two RCTs are ongoing, and we will present the trial design. We also discuss the operational aspects of incorporating experimentation in an innovation agency.

2 INTRODUCTION

Policy experimentation can be a useful tool in guiding innovation policy making by supporting more informed decisions in a complex area of policy (Bravo-Biosca, 2016). Experimentation is not unidimensional in its application, namely the pursuit of growing scientific knowledge, but can be deployed in various contexts with various objectives (Bravo-Biosca, 2020). It can be leveraged for exploratory and discovery purposes, such as to test a causal mechanism or assumptions about a problem, as well as the feasibility and potential of a new intervention (Ludwig, Kling, & Mullainathan, 2011). It can also be used to directly evaluate or optimize policy interventions – whether for measuring the actual impact of an intervention (e.g., impact assessment) or testing methods for policy intervention delivery or process improvement.

The objective of this paper is to highlight how Randomised Controlled Trials (RCTs) can be leveraged to supplement the evaluation of the impact of funding programmes and assist in the development of programme services and support measures in innovation agencies. In the next section, we present three Randomised Controlled Trials implemented in the Austrian Research Promotion Agency (FFG) used to evaluate the effectiveness of new measures intended to help strengthen R&I in start-ups and SMEs. Then, we discuss the learnings from these experiments as well as experimentation beyond an individual experiment. In the final section, we conclude.

3 RANDOMISED CONTROLLED TRIALS IN IN THE FFG

RCT is a specific form of research design that, in theory, goes beyond identifying correlation and instead provides more robust causal estimates. In principle, the fundamental design of an RCT is rather simple: participants in a sample (whether it be individuals or businesses) are randomly allocated to different groups, with each group receiving an intervention, and in the best case, one group receiving nothing (often called a control group). The impacts of the intervention(s) on specific outcomes are then compared across groups. The causal impact of the intervention can then be estimated while addressing potential selection bias, because the only difference between the groups, on average, is the randomisation. In practice, however, undertaking an RCT comes with some shortcomings; it is a rigid research design that requires a high degree of precision during the planning and implementation stages in order to acquire valid data and results (Deaton & Cartwright, 2018).

RCTs have been the standard in health policy (e.g. double-blind RCTs in clinical trials) for decades and are widely used in developmental policy, e.g. by the World Bank together with J-PAL, a research centre performing randomized impact evaluations of policy for poverty reduction. This research design has also been gaining traction in the field of innovation policy where it continues to show a lot of promise (Firpo & Phipps, 2019). One of the first trials where RCTs were applied in innovation policy was in the Netherlands in 2004 and 2005 to test the effectiveness of innovation vouchers (Cornet, Vroomen, & Van der Steeg, 2006). There is, however, still much to be learned about how and when to implement them for maximal benefit (Bravo-Biosca, 2020).

In a wider effort to move towards more informed decision-making, the FFG developed and implemented three RCTs investigating measures to foster innovation capacity of SMEs and start-ips with funding from the European Union's Horizon 2020 "Innovation in SMEs (INNOSUP)" programme.

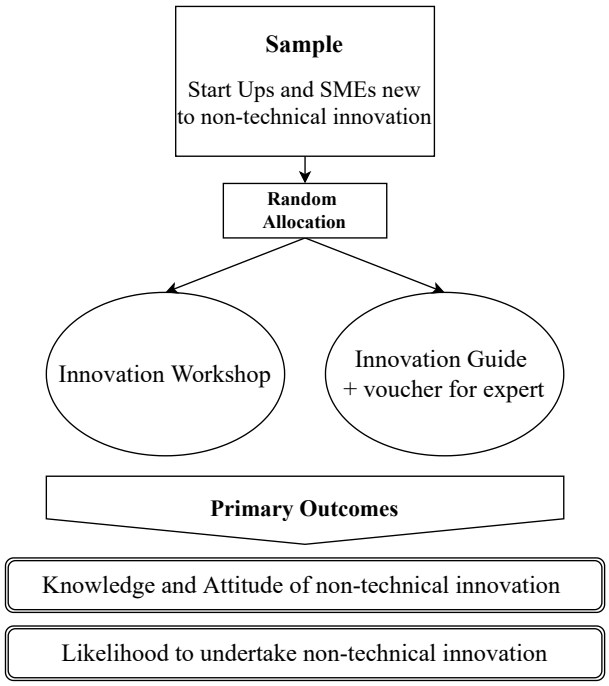
3.1 INNOVATION CAPACITY BUILDING IN SMES (INNOCAP, GRANT NR 824221)

Impact Innovation, a relatively new funding scheme in the General Programmes of the FFG (in German, "Basisprogramme"), was created with the purpose of providing a space for start-ups and SMEs to build know-how in non-technical innovation. Specifically, Impact Innovation funds early-stage, non-technical innovation projects with an emphasis on collaboration with users (or target group). Impact Innovation is designed to supplement traditional innovation policy measures, which primarily focus on advancing technical innovation and heavy investments in R&D, by bolstering the capacity of start-ups and SMEs to solve problems through "learning by doing", an approach that has been attributed with success in fostering innovation in European SMEs (Jensen et al., 2007; Parrilli et al., 2020).

An evaluation of Impact Innovation made evident that firms had shortcomings in planning and undertaking non-technical innovation projects. This was reflected in Impact Innovation proposals, where many applicants demonstrate a lack of understanding of iteration in project advancement, user involvement throughout the innovation process, and methods to manage and sustain innovation. Moreover, many firms who end up receiving funding in Impact Innovation still struggled with project implementation despite having a good project plan, citing a general lack of experience in innovation projects and associated methods. To address these shortcomings, two approaches to building knowledge on innovation processes were developed.

The RCT is therefore geared towards further developing Impact Innovation to better prepare innovation novices to undertake an innovation project. More specifically, we want to determine how best to improve firms' knowledge about and attitudes towards innovation processes thereby facilitating an improved implementation of innovation projects. To do this, the efficacy of two approaches on building fundamental knowledge of non-technical innovation process are tested.

Figure 1. InnoCAP Trial Diagram



SAMPLE

The target population are firms who are either considering starting innovation activities or already in the preparation phase of innovation projects. From the pilot RCT of InnoCAP, which tested measures to support innovation projects funded in Impact Innovation, we could infer that intervening after the project start (i.e., firms receiving funding) is too late for building up essential knowledge about innovation processes, as this knowledge could not be properly incorporated into the project plans. Thus, we targeted firms in earlier stages of their innovation projects.

We did not have direct contact with firms in the target population and thus had to recruit firms to the experiment. The recruitment process of our final sample consisted of two steps. First, firms were identified and contacted through two channels: Firms with an account on FFG's funding portal ("eCall") and firms in the FFG's multiplier network. The multiplier network consists of various innovation and entrepreneurial incubators, associations, and businesses that have a broad audience of start-ups and SMEs. Second, contacted firms were asked to fill in a survey, which measured baseline outcomes, relevant aspects for randomisation, and indicated actual interest in participating. The final sample comprised firms who completed the survey and is 61 firms.

A more detailed description of the recruitment process, including communication that was used, is available upon request. It will also be made publicly available with the final results.

INTERVENTIONS

Two interventions will be tested. The first intervention is an expert-led and peer-learning workshop in which experts from Impact Innovation provide input on essential aspects of the innovation process and experienced peers (previously funded Impact Innovation projects) share examples of how this might look in a project. In the workshop, firms will have the opportunity to discuss and work through their questions with both novice and experienced innovators in guided settings. The second intervention is a short guide highlighting important aspects of the innovation process and information on additional support resources, reflecting the content taught in the workshop. In addition, these firms will receive a voucher to an online expert platform, where they can link up with experts in a topic of their choice to assist them in development and implementation of a non-technical innovation project.

Due to the circumstances around firm recruitment to the RCT, it was not possible to include a control group in the experiment.

RANDOMISATION

The unit of randomisation was the firm, which were allocated evenly to one of two treatment arms. Randomisation was performed by researchers at the Vienna Center for Experimental Economics to minimize selection bias. Firms were block randomised on previous experience with Impact Innovation. In this context, previous experience with Impact Innovation is defined as having already received funding through Impact Innovation at least once. The rationale is that firms already funded by Impact Innovation have superior knowledge concerning innovation projects compared to firms with no prior experience with Impact Innovation. Eventually, 31 firms were allocated to the Innovation Guide and expert voucher arm (27 with no previous experience; 4 with previous experience) and 30 firms to the Innovation Workshop arm (26 with no previous experience; 4 with previous experience).

OUTCOMES

We are interested in three primary outcomes: *Knowledge of Innovation Process*, *Perceived Ability* and *Attitude towards Innovation*. *Knowledge of Innovation Process* assesses actual knowledge about the innovation process in Impact Innovation, specifically knowledge on problem-centered approach, iteration loops, innovation methods, and user involvement. *Perceived Ability* assesses an individual's perception of how well they can manage and implement the innovation process. *Attitude towards Innovation* assesses their perception of how advantageous they perceive non-technical innovation to be for firm development. All primary outcomes are measured with a pre- and post-intervention survey, measuring these indicators at baseline and shortly after the intervention period. We will be observing if there is improvement in actual knowledge, perceived ability, and attitude towards innovation projects.

As a secondary outcome, we are interested in *Innovation Activities*, specifically whether there has been an increase in innovation activities since the intervention. Six months after the intervention, participants receive another survey asking about the status of their innovation activities (e.g. applying for funding, working with experts, etc) since receiving the intervention. Furthermore, qualitative methods in form of interviews will be used to learn more about mechanics of the interventions.

The survey measuring the primary outcomes is available upon request.

RESULTS

The trial was run in the first quarter of 2022. We are still in the data collection phase. Results are expected by end of 2022.

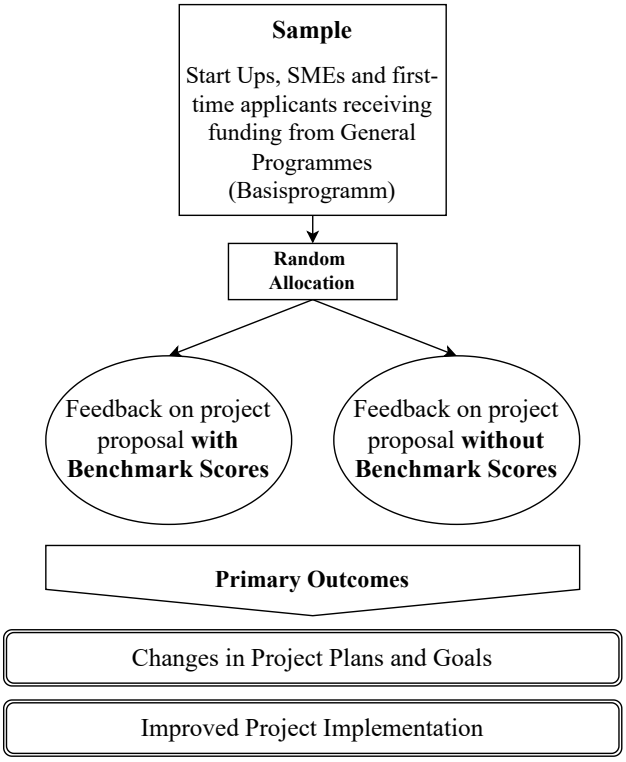
3.2 OPTIMIZING FEEDBACK FOR SMALL COMPANIES AND FIRST TIMERS (FEEDS FIRST, GRANT NR 824222)

The so-called "General Programme" is the largest and longest running innovation and research programme of FFG, open to all sizes of firms and types of research and innovation. An internal analysis of funding proposals in the "General Programme" found that the proposals of large enterprises who applied for the first time (subsequently referred to as "first time applicants"), SMEs and start-ups showed weaknesses in both the technical concept as well as the business plan. In order to provide additional support to SMEs, start-ups, and first time applicants funded in the "General Programme", we developed feedback on project proposals using data from the proposal evaluations outlining proposal-specific strengths and areas of improvement in four categories: Feasibility, Quality of the project, Utilization of project outcomes, and Sustainability.

Feedback given to SMEs and start-ups on business or innovation project proposals has generally been found to produce positive results (Wagner 2017). However, not all feedback is equivalent in evoking positive responses from those reading the feedback and initiating changes or improvements in business or innovation activities. While the information provided in the feedback is geared towards technical and business aspects of the project proposal at hand, people ultimately read, interpret, and integrate the feedback into the project. Thus, the presentation and type of information in the feedback as well as the person responsible for the project play an integral role in how and to what degree feedback is incorporated into the project (Mihm & Schlapp 2019; Yu, 2019; Liden & Mitchell, 1985).

In order to optimize the feedback on proposals funded by General Programmes, FFG and researchers at the University of Minnesota Carlson School of Management (UMN) developed an RCT to test the inclusion of relative ranking scores in feedback. The goal of the RCT is to understand if firms are more motivated to improve weak aspects of their projects when faced with their proposal's performance relative to their peers. Moreover, we will investigate the broader impact – negative or positive – of the intervention on the firm's likelihood to modify their project plan or project goals during implementation. We also hypothesize that outcomes may differ by size of organization (as categorized according to EU/FFG definition as either start-up, SME, or large enterprise).

Figure 2. FeedS First Trial Design



SAMPLE

The sample comprises 164 firms - 76 start-ups, 84 SMEs and 4 first time applicants – who received funding for innovation projects from the General Programmes in the period from November 2020 to December 2021. Our sampling procedure was a "trickle sample", as firms joined the experiment on a semi-rolling basis. The General Programmes has a permanent open call for proposals and makes a decision on the submission seven times annually, thus firms were inducted into the experiment in nine batches in line with the funding decisions.

INTERVENTION

The intervention was a relative ranking score for each of the four categories in the feedback. The relative ranking scores reflect the respective firm's performance in a specific category (e.g. feasibility) compared to all the firms in their batch (e.g. firm XY scored in the 60th - 80th percentile range in feasibility compared to other funded firms). The control group receives the same feedback sheet, only without the relative ranking scores.

RANDOMISATION

The unit of randomisation was the firm, which were allocated evenly to one of the two treatment arms. Randomisation was performed by research partners at the UMN in order to minimize bias. Firms were block randomized on firm size to ensure balance across both treatment arms. 84 firms were assigned to the control group comprising 39 start-ups, 42 SMEs, and 3 large organization. 80 firms were assigned to the treatment group comprising 37 start-ups, 42 SMEs, and one large organizations.

Due to the "trickle sample", we did not know the final sample size nor the composition of the sample in terms of firm size prior to randomisation. To accommodate this and ensure that balance was achieved, our research partners at the UMN generated a dummy sample using projections based on historical funding data to determine the composition of start-ups, SMEs, and first-time applicants. They then block randomized the dummies according to firm size. After every funding decision, firms "replaced" the next dummy in the corresponding block in the order in which they submitted their proposal, thus assuming the dummies' treatment arm assignment.

OUTCOMES

There are two primary outcomes of interest in this experiment, *Project Success* and *Project Changes*. *Project Success* is an ordinal variable assessing how well project implementation was executed. *Project Changes* is a binary variable measuring whether any deviations to the project plan (as set out in the proposal) occurred during implementation. Both outcomes are measured in a standardized procedure by the project evaluators at the end of the individual projects. Project evaluators were not aware of the treatment assignment of the firms whose projects they evaluated.

The secondary outcome of interest is *Perception of Feedback,* an ordinal variable where firms rate the utility of the feedback. We sent a survey asking the following question (translated from German to English) which was evaluated on a 4-point scale ranging from "Not at all helpful" to "Very helpful":

- Perception of Feedback
  How helpful did you find the Feedback?

Additional questions on specific aspects of the feedback (i.e. "What areas of feedback did you find most helpful? ") and open-ended questions on perceptions of feedback (i.e. "Please elaborate on how you incorporated the Feedback into the project.") were also included in the survey to get qualitative insights for the feedback, overall. The survey is available upon request.

RESULTS

Randomisation and induction of the experiment ended in December of 2021. Results are expected by end of 2022.

3.3 SOCIAL INNOVATION MATCHED CROWDFUNDING (SIM CROWD), GRANT NUMBER 824220

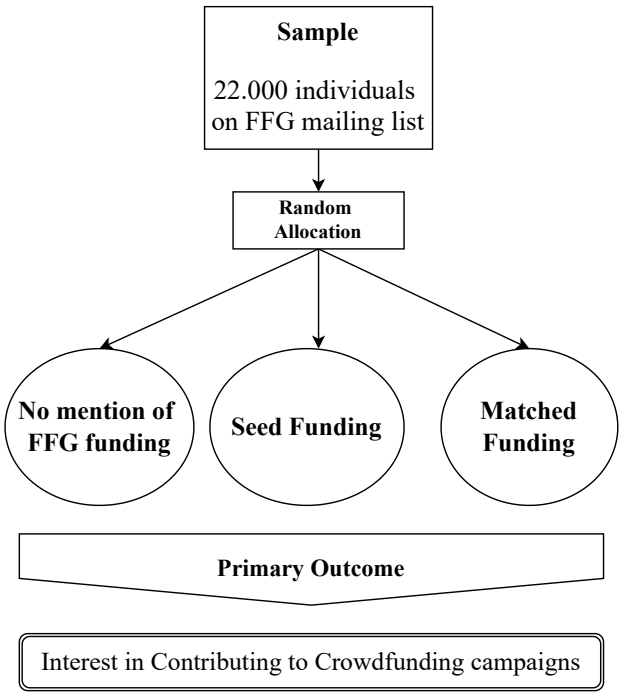
Social Crowdfunding, a new strand of the funding programme "Impact Innovation", was developed to provide social innovators and entrepreneurs an opportunity to acquire additional funding through crowdfunding (CF), because social innovators often face difficulties to meet the self-financing requirement for "Impact Innovation" funding. Crowdfunding is a viable source of alternative financing for SMEs. European SMEs collectively raised an estimated 22.3 billion euros in 2020 through

crowdfunding[1]. The premise of Social Crowdfunding is that social innovators receive support to develop and run an online, rewards-based CF campaign at the start of their innovation project, in addition to the funding received by "Impact Innovation". However, influencing individuals to financially support the project presents a hurdle for CF campaigns.

Online CF campaigns face the issue of information asymmetry given the distance between potential supporter and the organization running the CF campaign (Agrawal, Catalini & Goldfarb, 2014). To overcome this, organisations must communicate their credibility, relevance, and ability to deliver on the CF campaign goal as well as project goal through other means (Ahlers et al., 2015; Moysidou, 2017). This can be done by including high quality media, e.g. short video about the project, and showing positive comments about the project from other small contributors on the campaign platform (Courtney, Dutta, & Li, 2016), as well as publicizing financial support from major contributors (Karlan and List, 2020; Vesterlund, 2003). The type of financial support from major contributors matters, however, as different forms of support rouse different motivations of individuals to give or not. For example, seed funding may signal the quality of the project (List and Lucking-Reiley, 2002) while matched funding may give the impression of a reduced cost of giving (Karlan and List, 2007; Lee et al., 2017). Moreover, motivations to financially contribute to social initiatives differ among genders, often leading to different contribution patterns among males and females (Mesch, Brown, Moore, & Hayat, 2011; Mohammadi and Shafi, 2018). In the specific context of public grants and social innovations in businesses, the most advantageous form of support is still unclear, and there is little evidence on its impact.

To understand more about the efficacy of pairing FFG grants with crowdfunding campaigns for social innovations, FFG and researchers at UMN developed an RCT to investigate the impact of various forms of FFG support on individuals' likelihood to financially support a social CF campaign, specifically accounting for variations in gender (Bapna & Burtch, 2022). The final research design is a randomised, three-arm messaging trial testing whether the presence of public funding for a social innovation project has an effect on an individual's likelihood to contribute to a crowdfunding campaign for that project. As an additional proxy for quality signalling of FFG financial contribution, we investigated individuals' perceptions of the likelihood that the CF campaign and social innovation project will reach their goals. The differential impact of the treatments on all outcomes were also investigated with respect to gender.

Figure 3. Sim Crowd Trial Design



SAMPLE

Our final sample had 22.744 individuals. The sample comprises individuals in FFG's Customer Relationship Management (CRM) database who are interested in receiving emails about calls to apply for funding, workshop information, field-related updates, among other things. In terms of the gender distribution, individuals who identify as male made up a considerable portion of the sample, around 68% in total, whereas females made up around 32%. Nearly half (49%) of the sample is affiliated with Higher Education or a Research Institute while another third (34%) of the sample is working in the private sector. The remaining sample participants (19%) were split among other organisations or did not specify there employer. Regarding previous contact with the FFG, half the sample (49%) has had previous documented interaction[2] with the FFG of which 10.7% took place in the 12 months prior to randomisation.

1 Retrieved from Statistica (2021) https://www.statista.com/statistics/946659/global-crowdfunding-volume-worldwide-by-region/

2 Documented contact refers to in-person meetings, phone calls, emails, letters, or faxes that was noted in the CRM database

INTERVENTIONS

The interventions were a small additional text in the email detailing the nature of the FFG funding for the social innovation project in addition to information about the crowdfunding campaigns. The control group received no information regarding FFG funding of the projects; one intervention group received a text describing FFG support as Seed Funding; the other intervention received a text describing FFG support as Matched Funding. Below are the English translations of the texts.

i) Control – No mention of FFG funding

"We would like to introduce you to two projects that address social challenges and are currently seeking funding through a crowdfunding campaign. If you would like these projects to be implemented, you have the opportunity to support them."

ii) Seed Funding

Control Text +

„FFG has funded each project with the first 50% of the project costs, and the projects collect the remaining 50% through crowdfunding, among other means."

iii) Matched Funding

Control Text +

„FFG funds each project with 50% of the remaining project costs if the project secures the first 50% of the project costs via crowdfunding, among other means."

RANDOMISATION

The unit of randomisation was the individual, who were allocated evenly to one of the three treatment groups using a covariate constrained randomisation procedure. Covariate-constrained randomisation is a method used to ensure balance of baseline covariates among multiple treatment arms (Moulton 2004). Randomisation was performed by research partners at the UMN in order to minimize bias. As we hypothesized that outcomes would vary by gender, UMN block randomized on gender, while enforcing balance on additional available covariates such as affiliated organization and interaction with the FFG. Data on documented contact with the FFG, in particular, was included as a gauge for the likelihood that a recipient would even open the email. To see the output of randomisation and multinomial logit balance tests, refer to the publication of our research partners Bapna and Burtch (2021).

OUTCOMES

The primary outcome of interest was *Interest in Contributing*, a binary variable measured by whether an individual clicked on at least one link to visit the CF campaign webpage. The decision to click on the link and visit a CF campaign is an important precursor to actual funding (Bernstein et al. 2017, Bapna 2019, Bapna and Ganco 2021). Originally, we wanted to investigate the association between *Interest in Contributing* and actual contribution in this context, but it was not possible to measure the actual contribution of individuals directly, as there were many anonymous contributors and other difficulties in matching FFG data with CF campaign data.

In addition to *Interest in Contributing*, we investigated *Perceptions of Funding Risk* and *Perceptions of Project Risk* associated with different types of FFG involvement in the social innovation projects. Funding Risk refers to the likelihood that the crowdfunding campaign would (not) succeed in reaching its funding goal and Project Risk refers to the likelihood that the project itself would (not) be realized. We chose these outcomes because they are indicators of quality signalling from FFG involvement which are not directly linked to financial contribution from an individual, thus complementing *Interest in Contributing*. These outcomes were measured through a survey asking, among other things, these two questions (translated from German to English) which were evaluated on a 5-point scale ranging from "to a great extent" to "not at all":

- *Perceptions of Funding Risk*
  To what extent do you agree with the following statement: "Due to the involvement of the FFG, I see a lower risk for the projects to achieve their funding goals."

- *Perceptions of Project Risk*
  Assuming that the projects were able to achieve their funding goals. To what extent do you agree with the following statement: "Due to the involvement of FFG, I see a lower risk for the projects to achieve their project goals."

To ensure that respondents who answered these questions were aware of their respective treatment (i.e. those in the Control group were not aware of FFG Funding, and Seed and Match groups were aware of FFG funding), we asked a scanning question at the beginning of the survey:

"Were you aware that the FFG contributed financially to the projects Wochenplan.digital and mitwirken.at projects?"

To which respondents could answer Yes, No, or I do not know. Only responses to the survey where respondents in the Match and Seed condition responded 'yes' to the question, and respondents in the Control condition responded 'no' were included in the analysis.

RESULTS

A full analysis of the RCT was conducted by our research partners at the University of Minnesota and is available to read in Bapna and Burtch (2021). The results show that financial support from the FFG – in particular seed funding - signals quality of the social innovation projects and leads to a higher interest in contributing among females. When looking at the primary outcome - *Interest in Contributing* - females in Seed treatment arm demonstrated significantly more interest compared to females in the Control and Match treatment arms – 2.4 percentage points (50% relative increase) and 1.8 percentage points (35% relative increase), respectively. There was no statistically significant difference among men in different treatment groups, or in other words, men's *Interest in Contributing* was not deterred by FFG financial support. Furthermore, there is evidence that indicating financial support from the FFG in the form of Seed funding positively affected females' perception of campaign

success and project success, although these perceptions also varied by gender. When evaluating *Perceptions of Project Risk,* females in the Seed group perceived the social innovation projects to be significantly more likely to succeed in realizing their project goals compared to females in the other two treatment groups and males in the Seed group. There was no significant difference in *Perceptions of Funding Risk* among females in the three treatment arms.

4 LEARNINGS BEYOND INDIVIDUAL EXPERIMENTS

RCTs are not the only form of experiments that FFG conducts. Pilot actions[3], shadow experiments[4] or messaging trials (A/B tests) are examples of experiments that contain elements of RCTs, like randomisation or control groups, but are less strict in their design. What they all have in common is that they formulate ideas and hypotheses that are then tested or trialled in a structured and transparent way. They usually have a clear timeline, a concept how evidence will be collected and checkpoints at which results are assessed; and all this is designed *before* the experiment starts. An overview of such experimental approaches was published by Nesta (Hopkins, Breckon & Lawrence, 2020).

Getting to run a trial is often a process of constant negotiation and gentle nudges. A temptation might exist to jump head first into a large randomised controlled trial, however experimentation is not a cut-and-dry approach. Rather, running a trial is a process that may challenge the traditional ways of doing things and take an organisation into unknown territory. Designing experiments requires thinking "outside the box" of traditional processes and can have a positive impact on the discussion culture in innovation agencies. Starting small and bringing in external expertise also helped build up confidence, as trials do not always go according to plan. When they do not go according to plan and show unexpected results, they may trigger lively discussions, reflections, and a deeper understanding of underlying forces. This tends to open up "learning spaces" in which traditional standards and procedures can be discussed, and sometimes even modified, leading to more effective processes.

The development and implementation of experiments, in particular of RCTs, requires skills and an organizational culture not necessarily present in a public innovation agency. Engaging in experimentation therefore drives staff and skill development, however, it was also advantageous to collaborate with experts - FFG could tap into the valuable expertise provided by Innovation Growth Lab (IGL)[5] by Nesta and their extensive network of research partners, which, amongst other benefits, led to a collaboration with the University of Minnesota for SIM Crowd and FeedS First. If and to what extent an organisation embraces experimentation also depends on the appetite for risk. Experiments can "fail" in the sense that they show results that go against ingrained beliefs or traditional habits. To make the best out of such unexpected results, buy-in from senior management as well as an organisational culture that is prepared to accept and learn from such "failures" is very important. To this end, EU funding enabled us to pursue the RCTs described in this paper in a much more advanced form than would have been possible otherwise.

RCTs, in particular, are also a data heavy endeavour, which proved to be a double-edged sword. On the one hand, it gave us a structured way to exploit the mass amount of data already in the FFG, such as developing the feedback and benchmark scores in FeedS First, while at the same time innovating our data collection procedures by bringing new perspectives into the how, what, and when of data collection. On the other hand, (lack of) data was sometimes a limiting factor in being able to properly investigate a hypothesis. In certain cases, this was exacerbated by the EU GDPR regulations on data privacy, which made the process of data collection much more work intensive.

These positive effects come with a price. Experiments, especially full-fledged RCTs, are resource heavy, requiring both financial resources and time for a proper design, implementation and analysis. With the exception of messaging trials, they are better suited for bigger questions and where answers are not needed immediately.

5 CONCLUSION

RCTs and other forms of experimentation are an important tool in the set of methods used in evaluations as they open up new opportunities to test and evaluate the impact of programs and services. Evidence from SIM Crowd indicates that a seed funding commitment from FFG as a major donor can signal quality of social innovation projects. These results are informing future designs of the Social Crowdfunding program, particularly in how FFG grant should be communicated in CF Campaigns, but have also led to re-thinking future communications and the design of new funding measures beyond Social Crowdfunding. Although it is not yet possible to determine the impact on SME innovation capacity of the measures being tested in Inno CAP and FeedS First, carrying out the RCTs has already highlighted ways in which the FFG can improve data collection and external communication processes.

From an agency perspective, experiments create spaces for learning and change - elements necessary for a more modern innovation policy and agency. Experimentation moves the development of measures to capture and evaluate the impact of an intervention to the beginning of the experimentation process and strengthens the emphasis on developing clear impact pathways and indicators to measure success. At the same time, they also have a transformative effect on the culture as well as the way of working in an agency willing to engage in experimentation. However, experimentation can be resource intensive and demands specialized skill, thus it should be approached mindfully and is best done together with academics or other experts.

3 Pilot actions are used when the exact design of a new or changed programme or programme service is difficult to anticipate. Based on hypotheses and indicators defined at the beginning, the effect of the pilot action is measured and the design of the programme or service is adjusted accordingly.

4 Shadow experiments are typically data-driven experiments used to test variations e.g. in the way FFG reviews projects. These experiments are run parallel to the standard review process and do not influence its outcome but the results are used to optimize the standard processes. These experiments may or may not be randomised.

5 https://www.innovationgrowthlab.org/

6 REFERENCES

Agrawal A, Catalini C, & Goldfarb A (2014). Some simple economics of crowdfunding. Innovation policy and the economy, 14(1): 63-97. doi:10.1086/674021

Ahlers G.K., Cumming D, Günther C, & Schweizer D (2015). Signaling in equity crowdfunding. Entrepreneurship Theory and Practice, 39(4): 955—980. doi:10.1111/etap.12157

Bapna S (2019). Complementarity of signals in early-stage equity investment decisions: Evidence from a randomized field experiment. Management Science 65(2):933-952.

Bapna S & Ganco M (2021). Gender gaps in equity crowdfunding: Evidence from a randomized field experiment. Management Science 67(5):2679-2710.

Bapna S and Burtch G (2021). An Experimental Evaluation of Major Donor Funding Schemes for Crowdfunded Social Ventures. Available at SSRN: https://ssrn.com/abstract=3938577 or http://dx.doi.org/10.2139/ssrn.3938577

Bernstein S, Korteweg A, & Laws K (2017) Attracting earlyÐstage investors: Evidence from a randomized field experiment. The Journal of Finance 72(2):509-538.

Bravo-Biosca, A (2016). Experimental innovation and growth policy: Why do we need it? Innovation Growth Lab Policy Briefs, 1. https://innovationgrowthlab.org/policy-briefs

Bravo-Biosca, A (2020). Experimental Innovation Policy. In J. Lerner and S. Stern (eds.), Innovation Policy and the Economy, 191-232. Doi:10.1086/705644

Courtney C, Dutta S, & Li Y (2016). Resolving Information Asymmetry: Signaling, Endorsement, and Crowdfunding Success. Entrepreneurship Theory and Practice, 41(2): 265–290. doi:10.1111/etap.12267

Cornet M, Vroomen B, & Van der Steeg M (2006). Do innovation vouchers help SMEs to cross the bridge towards science? (Vol. 58). CPB Netherlands Bureau for Economic Policy Analysis.

Deaton A & Cartwright N (2018). Understanding and misunderstanding randomized controlled trials. Social Science and Medicine, 210: 2-21. Doi: 10.1016/j.socscimed.2017.12.005

Firpo T & Phipps J (2019). New Running experiments in innovation and growth policy: What can we learn from recent experience? FTeval Journal Research and Technology Policy Evaluation, 47(1): 46 – 50. doi: 10.22163/fteval.2019.328

Hopkins A, Breckon J & Lawrence J (2020) The Experimenter's Inventory: A catalogue of experiments for decision-makers and professionals. Alliance for Useful Evidence, Nesta.

Jensen M, Johnson B, Lorenz E, & Lundvall B.A. (2007). Forms of knowledge and modes of innovation. Research Policy, 36(5): 680-693. doi:10.1016/j.respol.2007.01.006

Karlan D. & List J.A. (2007). Does price matter in charitable giving? Evidence from a large-scale natural field experiment. American Economic Review 97(5): 1774-1793. doi: 10.1257/aer.97.5.1774

Karlan D. & List J.A. (2020). How can Bill and Melinda Gates increase other people's donations to fund public goods? Journal of Public Economics, 191(401):104296. doi: 10.1016/j.jpubeco.2020.104296

Lee D, Gopal A, & Lee D (2017) Micro-giving: On the use of mobile devices and monetary subsidies in charitable giving. Available at SSRN: https://ssrn.com/abstract=3280553 or http://dx.doi.org/10.2139/ssrn.3280553

Liden R & Mitchell T (1985). Reactions to Feedback: The Role of Attributions. Academy of Management Journal, 28(2):291 – 300. doi:10.5465/256202

List J.A. & Lucking-Reiley D (2002). The effects of seed money and refunds on charitable giving: Experimental evidence from a university capital campaign. Journal of political Economy 110(1):215-233. doi: 10.1086/324392

Ludwig J, Kling J.R. & Mullainathan S (2011). Mechanism Experiments and Policy Evaluations. Journal of Economic Perspectives, 25(3), 17-38. doi: 10.1257/jep.25.3.17

Mesch, D., Brown, M., Moore, Z., & Hayat, A. (2011). Gender differences in charitable giving. International Journal of Nonprofit and Voluntary Sector Marketing, 16(4), 342–355. 10.1002/nvsm.432

Mihm J & Schlapp J (2019). Sourcing Innovation: On Feedback in Contests. Management Science, 65(2):559-576.https://doi.org/10.1287/mnsc.2017.2955

Mohammadi A & Shafi K (2018). Gender differences in the contribution patterns of equity-crowdfunding investors. Small Business Economics, 50(2):275-287.

Moysidou K (2017). Motivations to Contribute Financially to Crowdfunding Projects. Open Innovation: Unveiling the Power of the Human Element, 283–318. doi:10.1142/9789813140851_0011

Parrilli M.D., Balavac M, & Radicic D, (2020). Business innovation modes and their impact on innovation outputs: regional variations and the nature of innovation across EU regions. Research Policy 49(8), 104047. doi.org/10.1016/j.respol.2020.104047

Vesterlund L (2003). The informational value of sequential fundraising. Journal of Public Economics, 87(3-4), 627-657. doi:10.1016/S0047-2727(01)00187-6

Wagner R.A. (2017). How Does Feedback Impact New Ventures? Fundraising in a Randomized Field Experiment. Available at SSRN: https://ssrn.com/abstract=2766566 or http://dx.doi.org/10.2139/ssrn.2766566

Yu S (2019). How does feedback Affect Entrepreneurial Performance? Available at SSRN: https://ssrn.com/abstract=3272525

AUTHORS

TESS LANDON
Austrian Research Promotion Agency (FFG)
Sensengasse 1, 1090 Vienna, Austria
E: Tess.landon@ffg.at
ORCID: 0000-0002-5491-994X

HARALD HOCHREITER
Austrian Research Promotion Agency (FFG)
Sensengasse 1, 1090 Vienna, Austria
E: Harald.hochreiter@ffg.at
ORCID: 0000-0003-1230-8486

KEYWORDS: Experimentation; RCT; SME support; innovation agency; innovation policy; public sector innovation

IMPACT ASSESSMENT OF BPIFRANCE'S FINANCIAL SUPPORT TO SMES' INNOVATION PROJECTS

MATTHIEU BRUN

DOI: 10.22163/fteval.2022.555

ABSTRACT

This paper evaluates the economic impact of Bpifrance's financial programmes to support SMEs' Research, Development and Innovation (RDI), called individual aid for innovation (IA). It focuses on the analysis of subsidies and zero-interest loans granted to SMEs over three years old during the period 2005-2018 in order to foster their RDI activity (R&D expenses and spending related to the development of innovative products, processes or services) and economic growth (turnover, employment).

We use a difference-in-differences methodology combined with a propensity score matching procedure to compare supported SMEs with non-supported SMEs with same initial characteristics. This counterfactual analysis is based on a unique dataset containing both financial and non-financial information about millions of French companies. Up to 12,000 SMEs supported over the 2005-2016 period have thus been analysed, making this study the first to estimate the effect of Bpifrance's individual aid for innovation on such a scale and using such detailed information.

Econometric results suggest that the use of Bpifrance's aid enables SMEs to boost their investment in RDI over the three years following the aid being granted, in comparison with non-supported SMEs with the same initial characteristics (+€250k of additional total R&D expenditure in aggregate over three years). Results show that the impact of the aid is additional, meaning that the support given does not take the place of any RDI investment that would have been made by SMEs had the support not been received. This analysis was supplemented by an examination of skilled employment and R&D employment, which indicates that Bpifrance's support encourages SMEs' investment in R&D jobs, with 0.5 more engineer/technician jobs per SME within three years (10% growth relative to the year preceding the support) and 0.4 more high-skilled jobs (up 9%).

Beyond the extra spending on R&D and innovation, the financial performance of supported SMEs also improves at the end of the three years compared with the counterfactual, with +€284k in additional turnover (6% higher than the year preceding the support), +€99k in additional added value, and +€77k more in export revenue. These effects are not significant at a one-year horizon, suggesting that the additional RDI investments resulting from the programme need time to be reflected in business growth.

1 ASSESSMENT OBJECTIVES

FOREWORD

This article is a simplified version of a report published in 20201 which formed part of an evaluation plan aiming at assessing the impact of a wide set of French public programmes dedicated to supporting innovation2, implemented upon request from the European Commission. The study was produced under the supervision of the French State and independent researchers from various institutions.

AN OPPORTUNITY TO ASSESS A BROAD PUBLIC PROGRAMME DEDICATED TO SUPPORTING INNOVATION

Companies possibly invest less in innovation than may be desirable for the whole economy. The existence and causes of the difficulties faced by companies in this respect have been widely documented in the academic literature. As regards the financing of innovation, some theoretical and empirical studies suggest that RDI projects may be particularly exposed to financial constraints (see Hall, 2002), notably due to their riskier nature (e.g., uncertainty about the commercial success of the product or service associated with the innovation project), and because RDI investments cannot be used to secure loans granted by private banks (RDI spending comprising mostly salaries). The technicality of innovation projects makes it more difficult for banks to assess the risk of such projects. In addition, innovation projects may intrinsically yield higher returns for the wider economy than for the individual companies that develop them because of the existence of positive externalities for example, or because

1 See https://www.entreprises.gouv.fr/files/files/etudes-et-statistiques/rapport_final_ai_bpi.pdf.

2 See https://www.entreprises.gouv.fr/fr/etudes-et-statistiques/autres-etudes/evaluation-des-aides-d-etat-la-rd-et-l-innovation-rapport.

companies may have difficulties protecting their inventions (thus not being able to benefit fully from their innovation efforts). This suggests that encouraging corporate investment in innovation through public intervention may be justified, particularly by mitigating the financial constraints innovative companies may face (either through the distribution of innovation grants, or through R&D tax credits for example). In this case, assessment is needed to determine the extent to which public money given to supported companies may have substituted for private innovation expenditures those companies would have incurred anyway (crowding-out effect) or whether the programme has had a positive effect on private innovation investment (crowding-in effect).

The empirical literature contains numerous studies assessing the effectiveness of public programmes aimed at supporting innovative companies through grants and subsidies, but very few focus on France[3]. Duguet (2004), using propensity score techniques, observes that the R&D subsidies granted by French ministries between 1985 and 1997 did not have a crowding-out effect on private R&D spending. Huber et Masquin (2012) show that the innovation projects supported by Oséo (Bpifrance's forerunner) are of relatively good technical quality since they are associated with a significantly higher production of patents than observed in non-supported companies. Serrano and Velarde (2009) find that the financial support granted by Anvar (Oséo's forerunner) may have partly crowded out some of the supported companies' private R&D expenditure (especially that of bigger companies). The present study aims to assess one of France's main innovation support programmes using a new and broad database containing precise information on SMEs over three years old[4], allowing evaluation of this programme on a very large scale. It attempts to measure the impact of Bpifrance's individual aid on the RDI spending of beneficiary SMEs and the extent to which the aid affected their economic growth in the following years (activity and employment).

2 BPIFRANCE'S INDIVIDUAL AID FOR INNOVATION

Bpifrance's individual aid for innovation combines seven mechanisms designed to finance RDI projects run by businesses individually, meeting financing requirements typically ranging from €30k to €200k. These projects are generally at an early stage of the innovation process, i.e.,

Table 1: List of innovation schemes related to Bpifrance's individual aid

Scheme	Companies / projects targeted	RDI expenditure covered by aid
Individual aid for RDI distributed by the Bpifrance network	SMEs and midcaps from all trade sectors	Industrial research and/or experimental development activities (building and developing prototypes, pre-production, pilot and demonstration installations, expenditure on intellectual property and standards compliance, market research, tests…)
French Tech grant	Start-ups (less than one-year old) with strong growth potential developing a business underpinned by an innovation	Internal or external costs directly linked to research concerning the design, scoping and feasibility of the project to be run (spending on support & guidance, intellectual property, feasibility studies, legal and market research, design, seeking partners, special training, travel, trade fair registration fees...)
Global Innovation Competition and the Innovation Competition	Companies with a disruptive economic model with the potential to grow internationally, and operating in specific fields of innovation (energy storage, plant proteins and plant chemistry, individualised medicine, collective security…)	Industrial research and/or experimental development activities
Fund for the Digital Society	Companies operating in the digital sector (such as nano-electronics, embedded software and smart objects, digital security…)	Payroll costs, R&D costs, acquiring patents, equipment and instruments used within the RDI project
i-Lab	Researchers seeking to create companies using their own innovations	R&D programmes needed to finalise the innovative product, process or technological service
Regional Innovation Partnerships	Non-technological innovation projects located outside Paris and its suburbs	Expenditure related to innovation's feasibility, development or production
Social Innovation Fund	SMEs and structures in the social and solidarity economy (non-profits and cooperatives) running innovative projects addressing a social need currently met poorly or not at all	Internal costs (staff assigned to the project, overheads and investment allocated to the programme), external costs (accommodation, support and consultancy services, feasibility studies, intellectual or industrial property rights and design services or specific training)

3 For the US, see Howell (2017) who, using a very robust methodology, finds that the SBIR subsidy programme had a substantial positive effect on small innovative firms' growth.

4 Data related to young SMEs (less than 3 years) are very scarce, so such firms may be largely underrepresented in our results.

before an innovation is likely to generate a potential economic benefit for the company producing it. All programmes are intended to finance RDI projects bearing uncertainty in terms of potential economic return for the company (see table 1):

- The individual aid for RDI distributed by the Bpifrance network[5] covers the majority of individual aid deployed in terms of both the amounts granted and the number of recipients supported;
- The French Tech grant is a mechanism specifically targeting start-ups;
- The Global Innovation Competition and the Innovation Competition are programmes targeting disruptive companies operating in specific fields of innovation;
- France's national Fund for the Digital Society is a wide-ranging programme combining various waves of competitions, all focused on the digital sector;
- The i-Lab competition targets researchers seeking to create companies using their own innovations;
- The Regional Innovation Partnerships and the Social Innovation Fund support non-technological innovation and social innovation projects located outside Paris and its suburbs.

Figures 1 and 2 present the evolution of the seven individual aid schemes in terms of both amounts granted and numbers of companies supported. As shown, individual aid distributed by Bpifrance's network is the oldest support mechanism for innovation used by Bpifrance. It is also Bpifrance's single most significant innovation support mechanism in terms of amounts granted and numbers of recipients (€400m in commitment and 2,600 recipients per annum on average over the last ten years). Consequently, the results presented in this paper will essentially cover this particular scheme. This programme is aimed at a broad target of eligible businesses (SMEs or midcaps[6] with no age or trade-sector re-

Figure 1: Numbers of recipients supported by individual aid
By year and mechanism
Source: Bpifrance

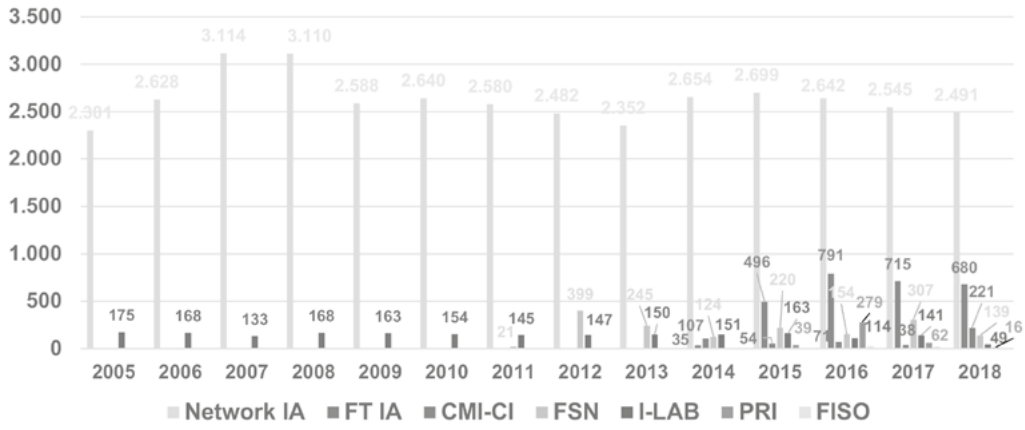
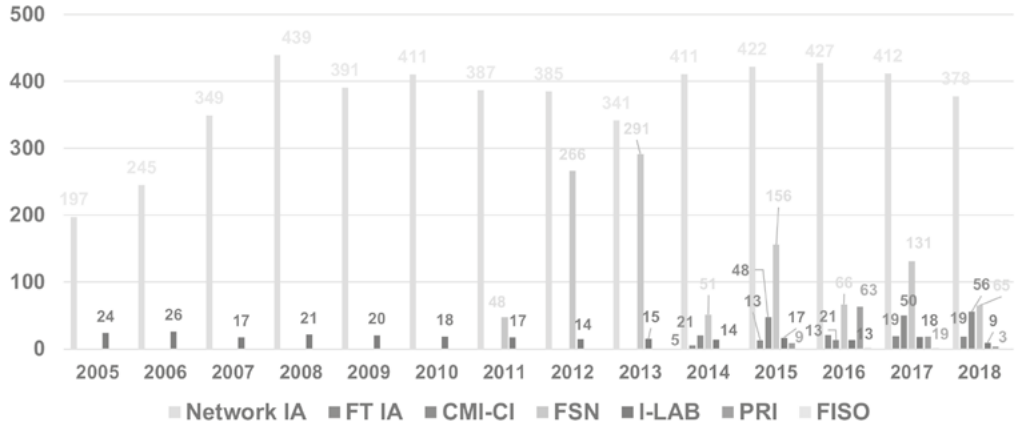


Figure 2: Amount of aid granted (€m)
By year and mechanism
Source: Bpifrance



5 This encompasses Bpifrance's network of regional branch offices, numbering around 50 in total in 2019, where the account managers specialising in examining innovation aid applications from local SMEs are based. Regional offices have decision-making authority up to a certain amount.

6 We define midcaps as companies between 250 and 4,999 employees at the group level.

strictions) and is intended to finance expenses directly linked to innovation development (industrial research and/or experimental development activities). Support takes the form of a subsidy, a repayable advance (subject to generating a certain turnover level) or an interest-free innovation loan (PTZI).

Network IA covers from 25% to 65% of the eligible expenditure basis, depending on the project and the size of the firm. The median amount for Network IA is €49k, and around 90% of the companies receiving Network IA are SMEs[7] (almost half with less than ten employees, see table 2). Over the 2005-2016 period, Network IA mainly covered manufacturing industry, information-telecommunications, and the scientific and technical activity sectors (see figure 3).

Figure 3: Breakdown of the number of recipients and amount of Network IA
By economic sector, period 2005-2016
Sources: Bpifrance, Ficus-Fare

Scope: Aid data for 2005-2016 where the French business registration number (Siren) and its trade sector are available.

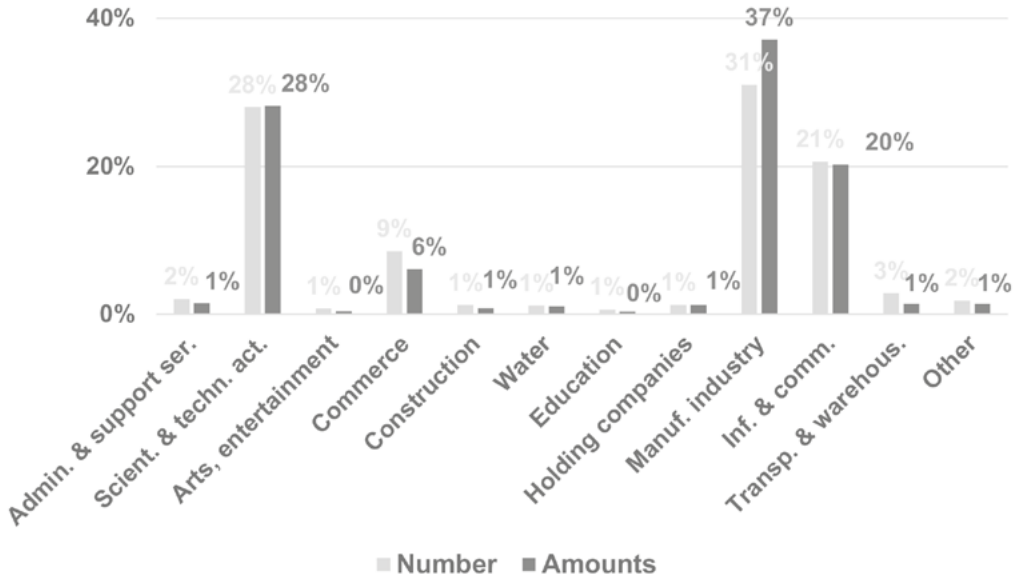


Table 2: Supported companies' financial statistics (Network IA)
Statistics computed the year the aid was granted, period 2005-2016

Network IA							
Indicator	**Observations** (number of firms x scheme x year)	**Of which, accounts available**	**Of which, data available** (share %)	**Mean**	**1st quartile**	**Median**	**3rd quartile**
Age	29.835	29.050	97%	13,7	2	6	16
Turnover	29.835	27.192	91%	6.651,0	132	843	4.060
Added value	29.835	27.192	91%	1.985,0	36	368	1.507
Headcount	29.835	26.894	90%	36,0	3	10	30
Capital expenditure	29.835	20.566	69%	215,0	1	17	100

Sources: Bpifrance, Ficus-Fare
Scope: Aid data for 2005-2016 where the French business registration number (Siren) is available.

7 Companies with less than 250 employees at the group level.

It is important to note that, at some point, companies benefiting from individual aid often use alternative public programmes intended to support innovation, such as the French R&D tax credit (CIR, generating nearly €7bn of support in 2018). They may also use a second tax credit scheme specifically designed to support young innovative businesses ("JEI", generating around €150m a year), or other direct grants or subsidies (See table 3).

Table 3: Numbers of Bpifrance individual aid recipients and their propensity to make use of other public support schemes for RDI Statistics computed the year the aid was granted, period 2005-2016

Year N	Number of recipients of Bpifrance IA in N	Of which having benefited from support between N-1 and N-3 (share %)				Of which having benefited from support in N (share %)			Of which having benefited from support between N+1 and N+3 (share %)*			
		Bpifrance IA	Other direct aid	Research tax credit	JEI	Other direct aid	Research tax credit	JEI	New Bpifrance IA between N+1 and N+3	Other direct aid	Research tax credit	JEI
2005	2.115					1%	na	19%	45%	2%	na	23%
2006	2.443					1%	na	21%	43%	2%	na	24%
2007	2.940					1%	na	19%	38%	4%	na	21%
2008	2.935	36%	1%	na	17%	1%	53%	20%	36%	5%	68%	22%
2009	2.481	33%	2%	na	17%	1%	59%	20%	36%	9%	71%	21%
2010	2.543	35%	2%	na	18%	3%	61%	23%	37%	10%	73%	24%
2011	2.513	35%	3%	57%	20%	4%	62%	24%	35%	12%	73%	26%
2012	2.817	31%	5%	60%	20%	6%	61%	24%	36%	14%	na	25%
2013	2.581	33%	5%	60%	21%	5%	64%	24%	35%	13%	na	28%
2014	2.823	30%	5%	55%	21%	5%	59%	26%				
2015	3.344	28%	5%	51%	20%	5%	na	24%				
2016	3.678	27%	4%	na	20%	4%	na	25%				
Total	33.213	32%	4%	57%	19%	3%	60%	22%	38%	8%	71%	24%

* Share in % among surviving firms 3 years after the aid being granted

Sources: Bpifrance, data from innovation support operators. "Other direct aid" covers Bpifrance aid to collective projects, and part of the aid from ANR (nuclear), ADEME (environment), CNES (space) and ONERA (aerospace) (source: France Stratégie).

Scope: All recipients of Bpifrance individual aid for innovation with a French business registration number.
Interpretation: In 2011, 2,513 beneficiaries received a Bpifrance IA, of which 35% had previously received at least one Bpifrance IA between 2008 and 2010.

3 IMPACT ASSESSMENT METHODOLOGY

DATA

Our first objective is to assess the individual aid's effect on supported companies' RDI spending, in comparison with the hypothetical situation of no support. A major obstacle in evaluating the aid's effect on RDI spending is the difficulty of reliably measuring such spending, which can cover wages (engineers, researchers…), prototyping and testing, market research and so on, none of which is easy to pinpoint in companies' financial statements as RDI spending. In this study, we measure RDI expenses in two ways:

- The GeCIR database (Research tax credit) is used to provide a first measure of the individual aid effect on companies' R&D spending. This database provides an annual list of companies that used the French Research tax credit scheme over the 2008-2014 period. It contains the amount of R&D expenditure such companies declared in order to benefit from the CIR tax credit, since this tax credit is computed as a proportion of their total yearly R&D expenditure. The GeCir database thus provides access to companies' R&D expenses, for companies that made use of CIR;
- DADS submissions (system for the automated reporting of employment data) are used to measure the effect of individual aid for innovation on R&D jobs, interpreting employment relating to technical roles within businesses as RDI labour. DADS provide accurate information on employment within any business that

employs staff, including data on the nature of the jobs, thereby making it possible to determine the numbers of engineers and technical staff employed and the high-skilled positions in a given legal unit, year by year. They are available for the period 1993-2016 and cover millions of businesses every year. This analysis supplements the examination of total R&D spending through the GeCIR database.

RDI investments are expected to affect the economic trajectory of the supported companies, through productivity gains, better market positioning, etc. If the individual aid has a positive impact on companies' total RDI spending, then its induced effects on recipients' economic growth should be isolated. We study supported companies' total turnover, export turnover and added value, as well as total recruitment, which are available in the Ficus-Fare tax statistics. This database contains complete economic and financial characteristics for almost all French businesses and covers the period 1994 to 2016.

Lastly, several databases enable companies' use of public innovation schemes to be tracked:

- The Bpifrance database enables companies using individual aid for RDI, the subject of this assessment, to be identified. These data cover the period 2005-2018 and around 25,000 distinct businesses;
- Companies' use of alternative RDI public aid programmes is identified through the following databases:
  - The GeCIR database described above enables companies using the tax credit scheme to be identified. As explained above, these data enable the performance variable to be built for measuring R&D spending, but they can also help to build the counterfactual sample used in the econometric analysis;
  - The JEI innovative start-up scheme database provides an annual list of businesses benefiting from this second tax scheme. Tracking this programme may be necessary since it is used a lot by companies that benefited from individual aid. Over the period 2004-2016, the GeCIR and JEI databases list approximately 50,000 distinct businesses;
  - France Stratégie gathered data related to various innovation aid operators in France, making it possible to build an aggregate variable identifying businesses receiving innovation grants other than the Bpifrance individual aid[8]. These data cover around 5,000 distinct businesses for all the operators combined.

In the longer term, the expected effect of RDI support policies also encompasses benefits for the community at large, benefits that are not necessarily monetisable by businesses (the spread of innovations into the rest of the economy, access to new healthcare methods, the reduction of pollution, etc.). These are positive externalities generated by innovation. It is difficult to measure the contribution made by an RDI aid programme to such collective benefits because they are influenced by many factors and are observed in the economy over a longer term. Such impacts are not studied in this assessment.

IDENTIFICATION STRATEGY

Our methodology relies on a difference-in-differences (DiD) approach combined with a propensity score matching procedure. The general idea of the DiD approach is to compare economic outputs of businesses that did and did not receive aid, on the assumption that in the absence of aid, their trajectories would have been similar or "parallel". The difference in the changes in performance metrics seen over time between supported and non-supported businesses is then attributed to the aid. To improve comparability between supported and non-supported eligible companies, we reduce the set of non-supported eligible companies to a subset of non-supported companies close to the supported businesses regarding different observable attributes in the past (counterfactual sample). These attributes need to influence both companies' likelihood of receiving an individual aid and their future performance. Counterfactual companies are chosen based on the following indicators:

- Use made of various innovation support mechanisms in the past is a factor expected to have a marked effect on the likelihood that aid will be used in the current year (see Duguet 2004): IA, other direct aid such as ADEME (environmental sector), ANR (nuclear), CNES (space), ONERA (aerospace), fiscal aid CIR and JEI;
- Ratios used to quantify companies' past innovation intensity level: the ratio between R&D spending and turnover in the year preceding receipt of aid[9], the ratio between the number of engineers and technical staff and the total headcount. The matching procedure also makes sure that the total amount of public-sector aid received by supported firms and non-supported firms over the last three years is similar;
- Qualitative factors (age, size, business sector, geographical location) and financial characteristics (past performance indicator level, growth in turnover and past capital expenditure, net profit margin, liquidity, added value over payroll costs, equity-to-assets ratio, debt coverage ratio, past gross operating profit margin) are also included in the propensity score matching.

Each supported company is matched with its closest non-supported counterpart based on the propensity score within a given individual aid cohort[10]. Econometric tests are run to ensure that the distribution of the above ex ante characteristics of both supported and non-supported companies are the same (balance tests). Additional econometric tests are run in order to check that the dynamics of the performance indicators of both supported and non-supported firms are similar before receipt of aid (falsification tests).

8 France Stratégie is a French think-tank that was a partner in this research work. The data supplied by France Stratégie have been gathered from ADEME (environmental sector), ANR (nuclear), CNES (space), ONERA (aerospace) and Bpifrance aid programmes for collective projects. They cannot be used here to pinpoint directly which type of direct aid a business used or the operator from which it was requested.

9 Such a ratio is available only for companies present in the GeCIR database.

10 The method used here is nearest neighbour with replacement. Supported and non-supported companies are matched only if sufficiently close in terms of propensity score, i.e., if the absolute difference in scores is less than a given limit. Various values for this threshold were tested with no significant impact on the results.

Once supported and non-supported companies are matched, we estimate the following model:

$$Y_{it}=T_{it}\,\delta+c_i+e_t+u_{it} \quad (1)$$

Where:
The index i means a company (supported with an IA or not)
The index t means the period
Y_{it} is the performance indicator
T_{it} =1 if the company i receives an IA in year t
c_i is an individual fixed effect
e_t is a time fixed effect

The coefficient δ measures the impact of aid on the performance indicator[11]. Performance indicators may not be available for very young businesses (less than three years old), which are then under-represented in the analysis. Moreover, impact estimates are made only on the SME population because the quality of matching for the midcap segment is very poor (see table 4). Results are therefore only valid for relatively mature SMEs[12].

Table 4: Analysis of matching quality: SMEs vs midcaps
Characteristics of companies on the matched sample

Sub-population	Average headcount N-1 (Nb)		Average total public aid granted between N-3 and N-1 (€k)	
	Non-supported	Supported	Non-supported	Supported
SMEs	17	17	79	80
Midcaps	253	1118	524	2412

18,344 observations used (N=year of receipt of aid)
Sources: Bpifrance, Ficus-Fare tax statistics, DADS employment data, data from innovation support operators

Scope: All businesses eligible for Bpifrance individual aid for which financial statements are available

4 RESULTS

MATCHING PROCEDURE

Estimation of the probability that a SME receives Bpifrance individual aid in a given year is in line with results previously found in the literature:

- Consistent with the work of Duguet (2004), the likelihood of using Bpifrance IA in a given year is strongly influenced by having previously received support for innovation in the last eight years:
  - This holds particularly true for Bpifrance individual aid received in the past, where the positive influence on using aid in the current year is substantial;
  - Likewise, use of research tax credits and having been part of the JEI innovative start-up scheme both have a strong positive influence on the likelihood of obtaining Bpifrance individual aid in a given year;
  - The proportion of engineers and technical staff in the total workforce in the year preceding receipt of aid is higher for supported companies than for non-supported ones;
  - The likelihood of using aid increases significantly with the intensity of R&D activities as measured by the past ratio of the total R&D spending as a percentage of turnover;
  - Other factors may also explain the probability of using an individual aid, notably:
    - Companies requesting Bpifrance individual aid have more frequently been exporters (in the past) than the rest of the companies eligible for aid;
    - They are significantly younger;
    - They are better-capitalised.

The latter results suggest that for our counterfactual analysis to be valid, supported SMEs need to be compared with innovative non-supported SMEs, which was expected[13].

11 When estimating the model, the distribution of the performance indicators is trimmed (1% to both the right and the left of the distribution) in order to remove the effect of outliers.

12 The analysis also excludes some trade sectors and forms of legal entity: limited partnerships, non-profits, the public sector, property development businesses, holding companies and the agricultural and financial sectors.

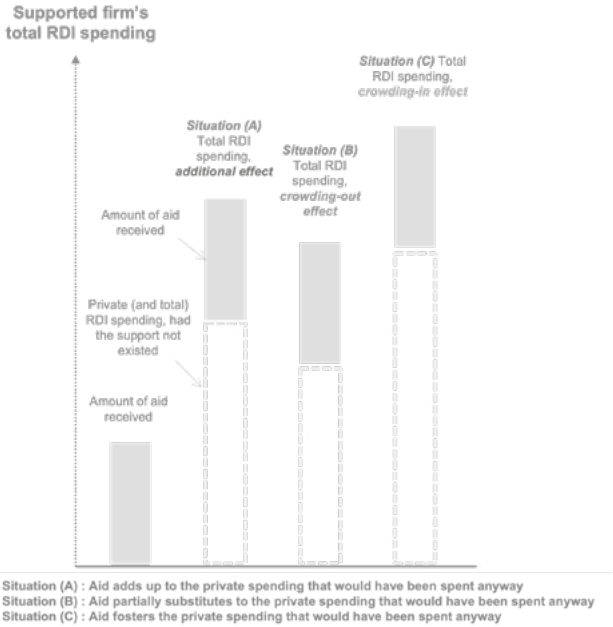
13 Econometric tests suggest that the hypotheses underlying the validity of the approach are verified. Eventually, supported and counterfactual non-supported SMEs are indeed very similar before receiving aid in terms of the characteristics mentioned above.

DISCUSSION OF THE METHODOLOGY AND ITS LIMITS

The validity of our methodology relies on the assumption that the performance of supported SMEs and their non-supported counterparts would have been close, had the aid not existed. This hypothesis raises the question of why the non-supported firms did not request access to the aid, since it would have helped them achieve better economic performance. In other words, our methodology may be valid if there are good reasons to believe that non-supported firms did not benefit from the aid because of specific factors unrelated to their future performance. A possible explanation may be that not all firms were aware of the existence of Bpifrance's programme or knew how to gain access to it (for example because such businesses were located far from Bpifrance's agencies or because the public schemes to support companies' innovation in France is known to be complex[14]), but this would need further investigation.

Moreover, our approach implies that at some point, we may compare currently supported SMEs with currently non-supported SMEs that received some Bpifrance aid in the past. Recent progress in the econometric literature show that under certain circumstances (typically when the effect of the programme varies with time or from company to company), estimations obtained using our method may be biased. Thus, further developments would be needed in order to assess the robustness of our results (see Baker, 2021).

Figure 4: Illustration of how the aid can interact with SMEs' private R&D expenditure
Additional effect vs crowding-out effect vs crowding-in effect



IMPACT OF BPIFRANCE'S INDIVIDUAL AID ON R&D SPENDING

We study the impact of Bpifrance individual aid on the total R&D expenditure reported by SMEs benefiting from the R&D tax credit. We also analyse the effect of aid on total R&D expenditure after deduction of any public support received by SMEs (private R&D spending). Figure 4 illustrates how to interpret the results of impact estimation, depending on how supported SMEs' private R&D expenditure evolves compared with the counterfactual situation. The impact of aid on private R&D spending can be negative, zero or positive: a negative impact means that at least part of the public aid granted to the SMEs was used as a substitute for private R&D spending (that would have been spent in R&D had the programme not existed). A zero-impact means that the public aid was entirely spent on R&D (in addition to SMEs' private spending). A positive impact means that the public aid encouraged SMEs to spend even more private funds in R&D than they would have invested if the programme had not existed.

Table 5 summarises the results of the econometric analyses: individual aid leads to an average additional increase of €36k in total R&D spending per firm in the year of the aid, and €250k in cumulative spending after three years, compared with the counterfactual situation. Access to Bpifrance aid permits SMEs to increase their total R&D spending compared with the counterfactual situation. Results also show a significant decrease (€74k) in privately sourced R&D spending in the year of aid, suggesting a crowding-out effect in the very short term. However, cumulatively over the three years following the support, individual aid has no impact on privately sourced spending (down €18k, result statistically not significant). Bpifrance IA finances projects spread over several years (typically up to three years), and it is possible that a time-lag effect exists between the payment of the aid and when it is actually spent on innovation projects, which might explain the negative impact on privately sourced spending in the very short term. Nonetheless, when the three-year assessment of this impact on total privately sourced R&D cumulative spending is produced, the effect of aid is additional, meaning that all aid disbursed was spent on R&D and privately sourced spending was unaffected.

14 See for example "Fifteen years of innovation policies in France" report from the National Commission on Innovation Policy Evaluation, France Stratégie, January 2016.

Table 5: Estimated additional impact of Bpifrance individual aid on R&D spending
Impact measured in €000s

Indicator	Supported SMEs studied	Indic. av. N-1 recipient	Estimated mean impact	95% CI low. limit	95% CI upp. limit	P-value	Relative mean impact
Cumulative R&D spending in the year of the aid (GeCIR)	3.889	387	36,1	28,2	43,9	0,000%	9%
Cumulative R&D spending over 3 years (GeCIR)	2.024	387	249,9	131,2	368,5	0,004%	65%
Cumulative R&D spending net of public-sector aid in the year of the aid (GeCIR)	4.116	209	-74,1	-81,2	-67,0	0,000%	-35%
Cumulative R&D spending net of public-sector aid over 3 years (GeCIR)	1.987	213	-18,2	-88,4	52,1	61,258%	-9%

Source: Bpifrance calculations
Scope: SMEs supported in 2009-2014with records in the GeCIR database for at least one year.

Note: Column 2 shows the number of supported SMEs analysed over the period, after matching and trimming of the performance indicator distribution. The mean impact and the limits of its 95% confidence interval are shown in columns 4 to 6. The relative mean impact is the mean estimated impact (column 4) relative to the mean of the performance indicator in the year preceding receipt of aid (column 3).

Table 6 offers a better understanding of the changes in privately sourced spending because the analysis focuses on a balanced panel of SMEs (those for which observation was possible for each of the three years following the support): the negative effect of aid on cumulative privately sourced spending declines over time, becoming not significant within two years. Such a result confirms that the crowding-out effect is only temporary.

Table 6: Estimated additional impact of Bpifrance individual aid on private R&D spending
Impact measured in €000s on a balanced panel, by year

Timeframe	Supported SMEs studied	Estimated mean impact	Standard deviation	T	P-value
Year of the aid	1.987	-52,2	7,1	-7,4	<0,01%
1 year after the aid	1.987	-40,2	14,0	-2,9	0,400%
2 years after the aid	1.987	-19,0	23,9	-0,8	45,620%
3 years after the aid	1.987	-18,2	35,8	-0,5	61,260%

Source: Bpifrance calculations
Scope: SMEs supported in 2009-2014 with records in the GeCIR database for at least one year.

IMPACT OF BPIFRANCE'S INDIVIDUAL AID ON R&D LABOUR

Table 7 suggests that use of Bpifrance individual aid results in a significant increase in the number of engineers and technical staff in recipient SMEs in comparison with the counterfactual situation, with an extra 0.4 jobs per SME in the year of the aid and 0.5 extra jobs after three years. For SMEs already employing people in these positions in the year preceding receipt of aid, no associated salary increase is observed, suggesting that aid is used to recruit new engineers and technicians, rather than to increase the pay of existing staff. It is interesting to note that the increase in R&D employment measured in this way is partly driven by SMEs employing engineers and technicians for the first time. The proportion of supported SMEs employing staff in this category before receiving aid climbs from 77% the year preceding aid (77% for similar but unsupported businesses) to 84% within the three-year timeframe (78% for the unsupported businesses). This suggests that Bpifrance IA encourages SMEs to invest in R&D jobs for the first time.

These results are corroborated by the analyses of highly skilled labour employment, which increases in similar proportions to that of engineers and technical staff in comparison with the counterfactual.

Table 7: Estimated additional impact of Bpifrance individual aid on R&D employment
Impact measured in headcount or €000s

Indicator	Supported SMEs studied	Indic. av. N-1 recipient	Estimated mean impact	95% CI low. limit	95% CI upp. limit	P-value	Relative mean impact
Engineers and technical staff, year of the aid	7.839	4	0,4	0,3	0,4	0,000%	9%
Payroll costs for engineers and technical staff, year of the aid	7.855	143	10,3	8,0	12,5	0,000%	7%
Engineers and technical staff after 3 years	3.718	5	0,5	0,3	0,6	0,000%	10%
Payroll costs for engineers and technical staff after 3 years	3.720	164	17,8	10,1	25,4	0,001%	11%
Average salary of engineers and technical staff, year of the aid	2.751	33	0,0	0,0	0,1	72,354%	0%
Average salary of engineers and technical staff after 3 years	1.576	35	0,0	0,0	0,1	51,186%	0%
Highly-skilled jobs, year of the aid	7.823	4	0,3	0,2	0,3	0,000%	7%
Payroll costs for highly-skilled jobs, year of the aid	7.823	190	11,5	9,0	14,1	0,000%	6%
Highly-skilled jobs after 3 years	3.724	4	0,4	0,2	0,6	0,002%	9%
Payroll costs for highly-skilled jobs after 3 years	3.729	212	18,4	9,1	27,7	0,010%	9%
Average salary of highly-skilled jobs, year of the aid	3.057	45	0,0	0,0	0,1	39,372%	0%
Average salary of highly-skilled jobs after 3 years	1.649	48	0,0	-0,1	0,0	43,722%	0%

Source: Bpifrance calculations
Scope: SMEs supported in 2009-2014 with records in the DADS database for at least one year.

IMPACT OF BPIFRANCE'S INDIVIDUAL AID ON EMPLOYMENT AND BUSINESS

From the point when aid is likely to affect SMEs' RDI investment favourably, it might be expected that such investment would result in a tangible effect on the economic trajectories of businesses in the short term. The estimated impact of aid on total employment supports the result obtained for R&D employment, so use of Bpifrance individual aid by SMEs results in the creation of an additional 0.7 jobs per SME in the year of receipt of the aid, and 1.6 jobs in three years, according to Ficus-Fare tax statistical data (Table 8). The results obtained using DADS employment data are qualitatively similar.

Table 8: Estimated additional impact of Bpifrance individual aid on total employment
Impact measured in headcount

Indicator	Supported SMEs studied	Indic. av. N-1 recipient	Estimated mean impact	95% CI low. limit	95% CI upp. limit	P-value	Relative mean impact
FARE total workforce, year of the aid	12.124	14	0,7	0,6	0,7	0,000%	5%
FARE total workforce after 3 years	7.908	16	1,6	1,4	1,9	0,000%	10%
DADS total workforce, year of the aid	7.000	13	0,6	0,5	0,7	0,000%	4%
DADS total workforce after 3 years	3.303	17	0,9	0,5	1,3	0,000%	5%

Source: Bpifrance calculations
Scope: SMEs supported in 2005-2016 (Ficus-Fare tax data) or 2010-2016 (DADS employment data).

Table 9 shows that the impact of aid on total turnover and added value is almost nil and barely significant in the year it is granted, but it is positive and significant over the three-year timeframe (€284k additional turnover in comparison with the counterfactual, making an average impact of +6%). These results suggest that RDI investment funded through aid needs time before its effects can be seen in economic terms.
Aid's impact on export turnover is also both positive and significant in the medium term (€77k over the three-year timeframe). In practice, the proportion of total turnover generated by exports for SMEs with a non-zero export turnover in the year preceding receipt of aid shows little difference between supported businesses and the counterfactual (remaining the same at three years for both populations, at a level close to 30%). However, the proportion of SMEs generating export turnover grows significantly among supported SMEs (from 55% in the year preceding aid to 64% in the three-year timeframe), whereas the change is slight for the counterfactual (55% to 57%). In this respect, individual aid appears to help trigger the international expansion of supported SMEs.

Table 9: Estimated additional impact of Bpifrance individual aid on recipients' turnover and added value
Impact measured in €000s

Indicator	Supported SMEs studied	Indic. av. N-1 recipient	Estimated mean impact	95% CI low. limit	95% CI upp. limit	P-value	Relative mean impact
Total turnover, year of the aid	9.814	4839	51,3	19,7	83,0	0,149%	1%
Total turnover after 3 years	6.592	5125	284,2	193,7	374,7	0,000%	6%
Added value, year of the aid	8.499	1933	-13,3	-27,7	1,1	6,973%	-1%
Added value after 3 years	6.101	1953	98,7	62,7	134,8	0,000%	5%
Export turnover, year of the aid	11.992	878	18,4	6,1	30,8	0,331%	2%
Export turnover after 3 years	7.464	1009	77,0	40,7	113,2	0,003%	8%

Source: Bpifrance calculations
Scope: SMEs supported in 2005-2016 with records in the Ficus-Fare tax data.

CONCLUSIONS

This paper uses a very rich database to conduct an unprecedented impact assessment of Bpifrance's individual aid for innovation. Econometric results suggest that such aid has a positive effect on SMEs' RDI investment, whether in terms of R&D expenditure or spending on R&D labour. Analysing SMEs that had used research tax credits before accessing the aid suggests that Bpifrance's individual aid lowers privately sourced R&D spending in the year of receipt of the aid, showing a very short-term crowding-out effect on these SMEs. However, cumulatively over the three-year timeframe, the individual aid has no impact on privately sourced R&D expenditure, suggesting that the aid received is allotted entirely to medium-term R&D expenditure (additional effect).

RDI investment made using aid results, within the three-year timeframe, in increased total employment and turnover for recipient SMEs. Their total and export turnovers also rise significantly, with aid contributing in particular to triggering international expansion for SMEs previously operating exclusively domestically.

5 BIBLIOGRAPHY

Baker A., Larcker D. and Wang C., 2021. How Much Should We Trust Staggered Difference-In-Differences Estimates?, Working paper 21-112, Harvard Business School

Brun M., 2020. Evaluation de l'impact des aides individuelles à l'innovation distribuées par Bpifrance, Working paper Bpifrance, https://www.entreprises.gouv.fr/files/files/etudes-et-statistiques/rapport_final_ai_bpi.pdf

Duguet E., 2004. Are R&D subsidies a substitute or a complement to privately funded R&D? Evidence from France using propensity score methods for non-experimental data, Revue d'économie politique 2004/2, Vol. 114 pp. 245-274

Hall B.H. (2002), The financing of research and development, Oxford Review of Economic Policy, 18(1):35-51

Howell S., 2017. Financing Innovation: Evidence from R&D Grants, American Economic Review 2017, 107(4): 1136-1164

Huber D. and Masquin B., 2012. Le financement de l'innovation par Oséo, Lettre Trésor-éco no. 102, May 2012

Serrano-Velarde N., 2008. How do firms react to R&D subsidies? Evidence from regression discontinuity design, Working document – European University Institute, version dated 24 November 2008

AUTHOR

MATTHIEU BRUN
Evaluation Unit of the Evaluation, Research and Prospective Department, Bpifrance
27-31 avenue du Général Leclerc 94700 Maisons-Alfort, FRANCE
E: matthieu.brun@bpifrance.fr

KEYWORDS: Innovation grants and subsidies, Research & Development & Innovation spending, difference-in-differences, propensity score matching

SYSTEMS ANALYSIS IN INNOVATION POLICY EVALUATION THE UNFULFILLED PROMISE

MICHAEL ROTHGANG AND BERNHARD LAGEMAN

DOI: 10.22163/fteval.2022.556

ABSTRACT

Our paper addresses the question, why systemic approaches have only played a modest role in impact evaluations of innovation and technology programmes so far and examines possible reasons for this shortcoming, as well as discussing solutions that could be offered to remedy the existing deficit. While the need for a systemic approach to evaluations has been stressed quite often, the methodological challenges and reasons for the lack of systemic evaluations in practice have to our knowledge not yet been addressed in a systematical manner. This contribution is conceptual in nature and based on a review of the research literature on the use of systemic approaches in evaluations of the impact of R&I policy programmes. The analysis shows that the use of systemic methods encounters both epistemological and institutional obstacles. Suggestions are made for the further development of the methodological repertoire by including suitable systemic approaches.

1 INTRODUCTION

Systems analysis, a child of the development of applied sciences around the Second World War, has experienced periods of boom and bust over the past seven decades (Barbrook-Johnson et al. 2021; Williams 2015). Today, in view of climate change and the development of global IT networks and artificial intelligence, the prevailing circumstances seem more favourable than ever for a broad acceptance of systems thinking and complexity research. The increased attention that is given to systems thinking[1] is not only the result of the insightful model-based analyses of environmental studies (Meadows 2008), but also serves as a vehicle for spreading awareness of globally coordinated efforts to curb atmospheric warming (e.g., Ison and Straw 2020).

Innovation systems research, which was particularly driven by Scandinavian authors (e.g., Chaminade et al. 2018; Edquist 2005; Lundvall 2010), has made a significant contribution to the development of innovation economics in the closing decades of the 20th century. It emerged mainly against the background of the insight into the fundamental shortcomings of the "linear model of the innovation process", which characterised the views of outstanding theorists in the 1940s and 1950s when innovation economics was still in its formative phase (e.g., Bush 1960 [1945]). The system concept, originally geared to the study of national innovation systems, was successively narrowed down to more limited areas of the economic sphere which form subsystems of the broader national innovation system: sectoral and regional innovation systems, technology fields, and industrial clusters. Moreover, in the context of the recent "transformative turn" in the innovation policy literature, systems approaches have gained a strong momentum (e.g., Borrás and Edler 2020; Schot and Steinmueller 2018).

Against this background, it is surprising that systems thinking has never taken firm root in the policy areas responsible for technology and innovation, i.e., has not been properly internalised by decision-makers. As Borrás and Edquist (2019: 40-42) observe, most actions of innovation policy in industrialised countries are based on the linear model. They offer the explanation that the linear model is capable of conveying the impression of a clear, easily comprehensible causal connection between policy measures and their intended effects. While the theoretical research on national innovation systems has found great resonance in innovation research, its practical consequences seem rather difficult to grasp and manage in the political process or to communicate to the public.

This paper discusses the extent to which systemic thinking has penetrated the practice of evaluating innovation policy programmes to date, and explores causes of what we perceive to be a relatively weak response in evaluation practice to systems and complexity theory and innovation systems research to date. We focus on the application of systems thinking to the evaluation of individual innovation programmes that fulfil elementary complexity criteria. In contrast, Borrás/Laatsit (2019) analyse diffusion of system thinking and systemic evaluation practices in the innovation policy field of the EU28 in total. The following discussion is strongly influenced by the experience gained by the authors in Germany. However, we assume that comparable experiences can also be confirmed by a closer analysis of the innovation policy-oriented evaluation practice of other continental European countries. We argue that

1 Systems analysis is inevitably linked to research into the system complexity of the systems under study - the central property of complex adaptive systems. This should be taken into account when we refer to "systems thinking" and "systems analysis" in the following. In the following, we focus on systems, because studies that claim to capture the complexity of the object of study in one way or another do not necessarily have to adopt a systems perspective on the object, i.e., they do not have to focus on its dynamic whole.

due to the flexibility and breadth of the relevant methodological tools, unused potentials of system-oriented evaluation research could be activated without critical additional expenditure of human and financial resources.

The paper proceeds as follows: In section 2, we take a look at the results of a Scopus query of the coverage of scientific papers on systems analysis and systems evaluation (section 2). In section 3, we address the diffusion process of systems thinking in evaluation research in general as well as in innovation policy and its evaluation. Subsequently, complex innovation programmes are introduced as an object of investigation (section 4) and complexity attributes are demonstrated at the example of two German programmes (section 5). In section 6, we ask, what a systems evaluation is and address possible reasons for the low degree of reception of systems thinking in the evaluations in section 7. Section 8 concludes with practical suggestions for a pragmatic handling of systems approaches.

2 SYSTEMS ANALYSIS AND SYSTEMS EVALUATION PAPERS IN SCOPUS

The systems concept received considerable attention in innovation research, but it never became the dominant paradigm. The ground for a more systemically oriented evaluation practice in the field of innovation policy was therefore rather rocky in the beginning. In this context, the first question is to what extent systems thinking has influenced evaluation theory and evaluation research practice in general.

If systems thinking were to be used to a greater extent in evaluations of innovation policy programmes, this should have made itself felt in the relevant contributions to scientific journals. However, it should be noted that evaluations, which in most cases are commissioned research, do not necessarily result in journal articles. Nevertheless, a considerable and, it can be assumed, also representative part of practical evaluation research sooner or later finds its way into the world of scientific journals.

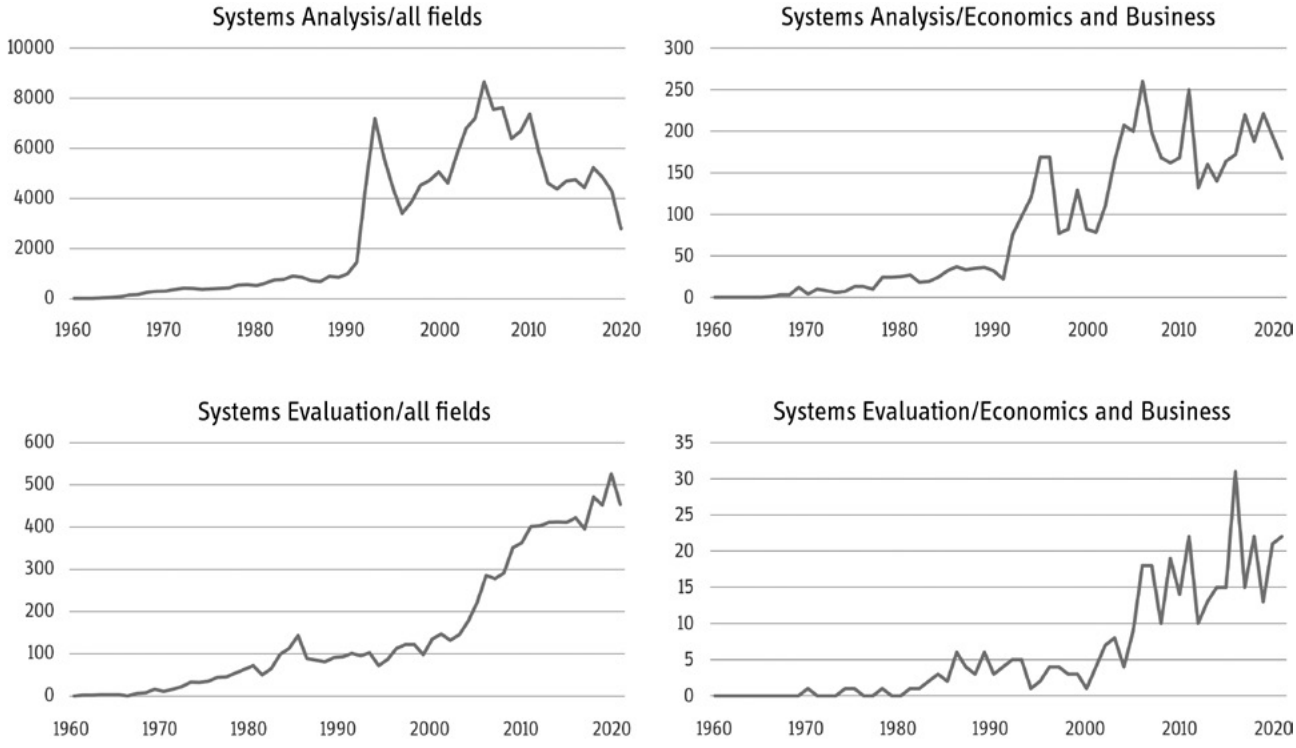
A recent search of the scientific bibliographic database Scopus showed that systems analysis, beginning in the 1960s, has found its way into the scientific literature. A strong upswing has been observed since the early 1990s, which reached a first peak in 1994 with 7.197 registered titles and its highest level to date in 2006 with 8.635 mentions (Figure 1, upper left side).

If one narrows down the search to the fields of "economics" and "business", comparatively low numbers emerge. For the period from 1960 to 2021, the highest value was reached in 2006 with 260 titles (Figure 1, upper right side). This corresponded to a share of 3.0% of all titles listed in connection with systems analysis. The highest share value was registered in 1978 with 5.6% (24 titles). A look at the predominant keywords of the registered titles shows that they are predominantly assigned to the fields of business administration/management, operations research and IT development.

In addition, the results of a Scopus search for titles containing the keywords "system(s) evaluation" or "systemic evaluation" are given in Figure 1 (lower side). The number of titles has increased significantly since the turn of the millennium. An analysis of the keywords shows a

Figure 1: Systems analysis and systems evaluation in the research literature as recorded in the database Scopus, all disciplines and economics/business, number of registered publications, 1960-2021

Source: Own data search and depiction; keywords "system analysis"/"systems analysis"; "system evaluation"/"systemic evaluation" in all areas and in the areas "economics" and "business" in the fields "title", "abstract", "keywords", 26 October 2021.



clear dominance of works that can be assigned to computer science in the broadest sense. As Figure 1 (lower right side) shows, the number of titles from the fields of "economics" and "business" is rather small, but it also increased substantially in the past decades. It reached its peak in 2016 with 31 registered works (7.4 % of all publications on "system(s) evaluation").

Our attempt to identify articles with the keywords "systemic evaluation" or "systems analysis" and "innovation policy" ended up with a short list of 9 titles in total for the entire period under review. The central result is, in other words, that there is no evidence for a more widespread discussion of systems evaluation in scientific papers in the field of innovation policy. We assess the relevance of this observation by taking a look at the general uptake of systems thinking on the one hand in policy evaluations in general and on the other hand in the field of innovation policy evaluations in the following section.

3 SYSTEMS THINKING IN (INNOVATION) POLICY EVALUATION

Theoretically oriented evaluation research has taken up the impulses of systems theory and its sibling complexity theory since at least the early 2000s. In the meantime, both streams of theory development have also found their way into practical evaluation work (Midgeley 2006). Admittedly, this applies more to the Anglo-Saxon countries than to continental Europe. The pioneering work of the Santa Fe Institute, New Mexico, founded in 1984 and the recent evaluation-oriented research of the UK Centre for the Evaluation of Complexity across the Nexus (CECAN) as well as the inclusion of complexity-related evaluation approaches into the 2020 edition of the Magenta Book (HM Treasury 2020, for the work of CECAN see the special issue of Evaluation, 1/2021, dedicated to the topic of complexity) are outstanding examples of the increased attention to systems approaches in evaluation research. Another milestone in the dissemination of systems thinking in evaluation research is the "Expert Anthology", published in 2006 by the American Evaluation Association (Williams and Imam 2006).

Evaluation research, as a branch of applied knowledge that is constantly striving to absorb new concepts, should actually have played a role as a pioneer of systems thinking in politics. Indeed, such a reception of systems theory thinking can be observed in evaluation research, albeit rather late. When the CECAN authors talk about a "turn towards complexity" (Barbrook-Johnson et al. 2021: 5; earlier use of the term in Mowles 2014), this applies not only to theoretical research on policy evaluation tasks, but also to parts of the practical evaluation activities in the Anglo-Saxon countries. However, even here, policy evaluations are much more focused on health care and CECAN's "nexus", namely food production, energy production, water management and the handling of environmental issues, than on the area of interest to us here, innovation policy. In contrast, a recent policy paper of the American Evaluation Association's Research, Technology and Development Evaluation Group (AEA 2015), to cite a current example, does not specifically address the possible use of a systemic perspective in the evaluation of R&D programmes.

In particular, those policy areas in which the problem pressure resulting from the perception of the complexity of the tasks to be solved was particularly substantial proved to be receptive to systems thinking. Prominent examples are environmental, health and development policy. The evaluation field in which the system perspective has probably gained the strongest foothold so far is the evaluation of development projects (e.g., Williams 2015 seems especially inspired by this field). This can be explained by the fact that in no other policy field is the pressure for a close examination of the impact of projects as great as in the field of development aid. After all, there are "lost decades" of many failed development aid projects in the poorest countries of the world, in which not only misconceived large-scale projects, but also many well-conceived manageable endeavours have proven to be failures in retrospect (Easterly 2007; Moyo 2009). The increasing insistence on conducting rigorous evaluations, preferably experimental designs and Randomized Control Trials (RCTs), where they are possible, as well as the strong emphasis on introducing a systemic perspective into the evaluation of projects, can be seen as a reaction to earlier failures.

But how is the present situation in the field of innovation policy evaluation? Unfortunately, there is no database on evaluation studies that could hold a candle to Scopus. The evaluation study database SIPER could possibly provide valuable information on the extent to which systemic thinking has found its way into the practical evaluation of innovation policy measures. Unfortunately, the search criteria do not yet allow external users to search for relevant keywords. A cursory review of evaluation studies recorded in this database shows at least that many evaluation designs contain methodological components that could also be part of system perspective evaluations. However, as will be shown below, this alone does not constitute a system evaluation.

This also applies to evaluations in the national framework of Germany, where the authors have practical experience in the evaluation business: We are not aware of any evaluation of a relevant programme that has explicitly committed to a systemic evaluation approach, although this would of course always be theoretically possible within the framework of a mixed method design.

At the same time, however, there is a certain unease among researchers. It can by no means be said that evaluation researchers are oblivious to the actual complexity of their objects of investigation when working on evaluation assignments. Researchers who have been in the field of evaluating complex state interventions are familiar with the gut feelings that structural and process characteristics of the impact patterns of innovation programmes are actually much more complicated than they appear in the indicators and methods used. Practitioners can only point out in their studies that there are still many relevant influencing factors and impact mechanisms that elude analytical access for the time being.

Innovation researchers have also recently repeatedly called for systemic evaluations of innovation policy. Edler/Fagerberg (2017) emphasise that "the available evidence on innovation policy impacts at the national level seems to suggest that holistic – or systemic perspectives in policy is important" (p. 14). The authors have primarily the call for cross-instrumental evaluations in mind when they emphasise that the overwhelming majority of evaluations to date have focused on a single instrument (2017: 13). In their comprehensive study of evaluation practices in EU28 Borrás/Laatsvit (2019) found that few countries have developed a system-oriented type of innovation policy evaluation. However, the lack of systemic analysis also applies to individual programmes that address a complex object of investigation such as clusters. As Uyarra/Ramlogan (2017) point out, the research methods used today are hardly capable of

adequately capturing the complex interactions of a multitude of actors who dominate this policy field.

4 COMPLEX INNOVATION PROGRAMMES AS OBJECT OF INVESTIGATION

It would hardly be possible to provide a binding uniform definition of the concept of system in view of the ubiquitous presence of dynamic systems in all areas of reality, and such a definition does not exist (so for example also Williams and Hummelbrunner 2011: 16). The meticulous, comprehensive attempt made by Ackoff (1971) half a century ago to define the system concept as precisely as possible is hardly suitable for capturing the diversity of what is rubricated under "systems research" today.

Essential elements of a pragmatic definition for practical use, however, that can be found throughout the contributions of the various authors are: (i.e., Ackoff 1971; Mainzer 2008, 2015; Meadows 2008): A system is a set of interconnected elements from any realm of physical or virtual reality that form a whole, which are in mutual dynamic relationships with each other and in their interaction can produce properties of the whole that are not inherent to the individual elements. This whole - the system - exhibits certain changing patterns of behaviour (roles or functions) that determine its interactions with its constantly changing environment.

The central property of the type of system we are interested in here – complex adaptive systems - is their increasing complexity over time, which guarantees their survival in a dynamic world requiring their balance between order and chaos (on complexity cf. Holland 2014; Mainzer 2008; Mitchell 2009). Therefore, the topic we are interested in here is analysed on an abstract basis both in systems theory and the related field of complexity theory. Complex adaptive systems are characterised by (e.g., Forss and Schwartz 2917; HM Treasury 2020; Mainzer 2008, 2015; Meadows 2008):

- the continuous adaptation to challenges from the system's environment or internal relations,
- the occurrence of feedback loops in the development processes of the system,
- the appearance of non-linearities in the development of the system elements and the system,
- the ability of the emergence of new properties at higher levels within the system,
- the evolution of the system in a self-organized way,
- the existence of (non-deterministic) path-dependencies in the evolution of the system, and
- uncertainty and resulting limited calculability of the development of the system including the generation of unexpected effects.

The innovation system as a whole and components of it are complex adaptive systems. We denote all programmes that are intended to influence the evolution of the innovation system as a whole or of one of its components as "complex innovation programmes". These components can be, for example, sectoral innovation systems, technology fields, technology clusters, or innovation networks.

The complexity of the matter addressed is usually reflected in the complexity of the programme, for example in complex target bundles that may themselves contain trade-offs between individual targets. For the classification of a programme as "systemic-oriented" or "complex", it is not necessarily decisive whether the programme makers are fully aware of the complexity of the programme object.

5 SYSTEMIC CHARACTERISTICS OF TWO KINDS OF PROGRAMMES

A systemic perspective of an evaluation makes sense if the programme under investigation is directed at an area of intervention that has systemic qualities and the intervention thus addresses systemic goals. The criterion for introducing a systemic perspective cannot therefore be, as Imam et al. (2006) correctly state, whether a programme is financially voluminous or not, or whether the implementation process is complicated. In the following, we examine two examples from federal German innovation policy to demonstrate that systemic approaches can be implemented, on the one hand, in large programmes that are well equipped with financial resources and, on the other hand, in the case of small programmes that are provided with little funding. Both programmes mentioned here belong to a group of programmes that can be classified as systemic instruments "avant la lettre" according to the analysis by Smits/Kuhlmann (2004).

Table 1: Characteristics of two German innovation policy programmes at the federal level

	Industrial Collective Research (IGF)	"go-cluster"
Funding Ministry	Federal Ministry of Economic Affairs and Energy (BMWi)	Federal Ministry of Economic Affairs and Energy (BMWi)
Year/Date of establishment	1954	1 July 2012
Financial scope of the subsidy	169 Mio. € in 2018 for around 550 R&D projects; total sum approx. 4 bn € cumulatively since the year of establishment (2020 prices)	1 July 2012 – 30 June 2015: 1.5 Mill €; in 2018 626 vouchers for the improvement of innovation management in participating clusters; total sum spent in 2012-2020 amounts to approx. 4.75 Mill €
General promotion objective	Strengthening the research base of medium-sized industry	Providing a stimulus to improve cluster management of cluster initiatives that meet certain quality criteria to enable them to turn in highly effective international clusters
Mediator organisation(s)	AiF – German Federation of Industrial Research Associations & 101 Industrial Research Associations	VDI/VDE Innovation + Technik GmbH as project executing agency; Internet presence at "Clusterplattform Deutschland"
Grant recipients	Research organizations (institutes of research associations, university institutes, other institutes)	Participating cluster initiatives (at present 84)
Use of the funding for	Funding of industrial research and development projects carried out by independent research institutes or university institutes; promoted R&D projects should address the research needs of SMEs in particular	Advice and training for cluster managers; support of knowledge exchange between national and foreign cluster initiatives; support in the establishment of international contacts
Sectoral and technological orientation	Open to all technologies and industrial sectors; traditional focus more on highly developed conventional technologies, recently increased presence of high-tech sectors such as the IT sectors	Open to all technologies and industrial sectors; innovative industrial clusters are to be promoted above all, so the actual preference is more for new technologies

Source: Own depiction, data on "go-cluster" for 2012 – 2015 from Eckert et al. 2016: 76; data for 2018 from Deutscher Bundestag 2019: 4-5; for detailed information about the IGF programme cf. RWI and WSF 2010.

The more comprehensive of the two programmes is the programme for the promotion of industrial collective research (IGF), while the "go-cluster" programme is very modestly funded. Both address structural policy objectives and are intended to strengthen the competitiveness and innovative strength of the German economy. Table 1 provides an overview of the major parameters of the two programmes.

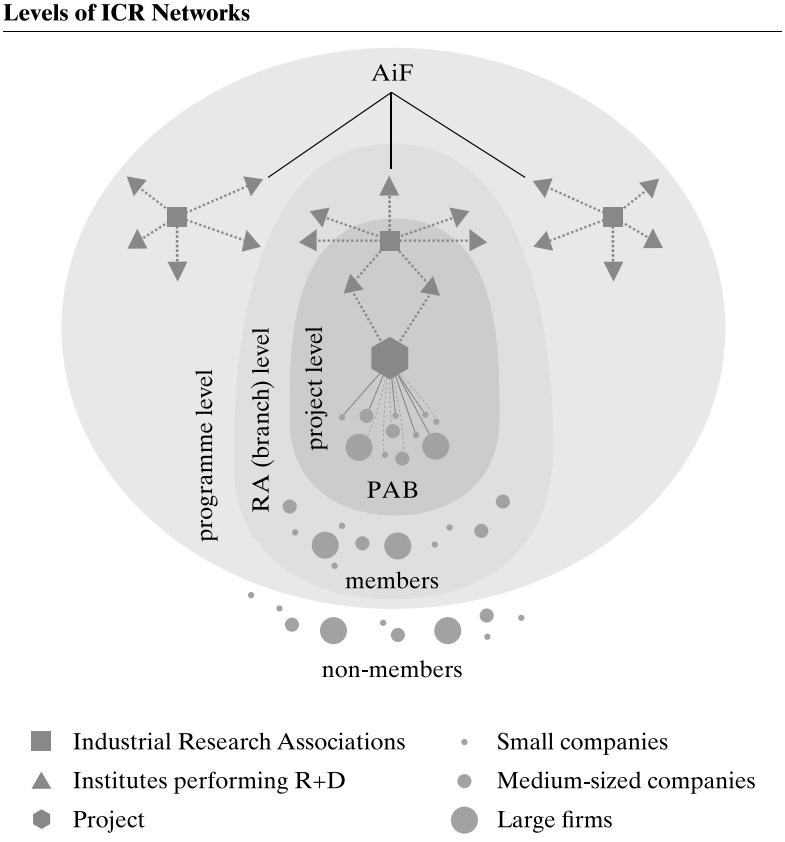
While the IGF focuses on funding collaborative research projects that serve specific research interests of companies and are often initiated by them in dialogue with research organisations (Figure 2), the go-cluster focuses on funding selected activities of the cluster management of "innovative clusters". In the case of IGF, the project proposals of the research associations are reviewed in a system-internal peer review process. In the case of go-cluster, the project management organisation reviews the applications of cluster initiatives that apply for participation based on an agreed catalogue of criteria. Successful participants are awarded one of the honorary labels (gold, silver, bronze) of the European Cluster Excellence Initiative (ECEI) in an internal selection process.

Both programmes aim to contribute to securing the long-term competitiveness of the German economy and thus, at least indirectly, to foster productivity, growth and job security. In both cases, the programme's aspirations go far beyond the immediate funding purpose and, in accordance with the programme logic, focus on central aspects of the long-term development of social welfare, although only in the case of the IGF these ambitions are backed up with substantial financial resources. In principle, both programme rationales can draw on the scientific authority of innovation economics research as well as cluster research. However, whether the programmes are actually suitable for producing the hoped-for positive effects on innovation and growth must - as always in such cases - be left to the results of evaluation studies.

What makes these programmes systemic? The answer to this question is to be found in the object of state intervention, in the actors on whose actions the success of the programmes ultimately depends, in the processes envisaged and the results aimed at:

- **Object of state intervention**
  According to the programme organisers (AiF), the IGF project funding is aimed at around 50,000 small and medium-sized enterprises and currently 101 research associations and research organisations are involved in industrial research. The IGF thus addresses an industrial research network that encompasses large parts of German industry. This comprehensive network consists of a multitude of nested individual networks (Figure 2) such as 101 industry level innovation networks[2] and hundreds of project level networks.
  Go-cluster (currently) supports 84 cluster initiatives that claim to represent a spatially located cluster consisting of vertically and horizontally connected companies, research organisations, and associated organisations.
- **Actors**
  In the case of both programmes, a large number of individual companies and research institutes are involved, whose development depends on a large number of individual decisions and internal and - only to a small extent influenceable - external influences. Thereby, the influence of the programme on the development of the individual firm in the vast majority of cases can only be of a marginal, hardly measurable dimension.
- **Processes**
  The development processes of the objects to be promoted are of an extremely diverse and complex nature and are in principle beyond the control of the state. They are characterised by feedback loops and are generally non-linear in nature. Emergence plays a role in the development of networks and clusters that are fostered by both programmes.
- **Results**
  The ultimate results of the complex state intervention, which occur over a long chain of indirect effects, are uncertain. They can neither be planned nor controlled in advance. In principle, considerable time elapses before the results of such an intervention can be realised. In the analytical identification of programme effects, an attribution problem has to be solved, as they are the result of the influence of multiple interdependent factors.

Figure 2: German Industrial Collective Research as system of systems



ICR – Industrial Collective Research (IGF)
AiF – Arbeitsgemeinschaft industrieller Forschungsvereinigungen (German Federation of Industrial Research Associations)
PAB – Project Advisory Board (Projektbeirat, members are firm representatives, accompanies individual research project)
RA – Research association (organizing pre-competitive research at sectoralindustry level)
Source: Rothgang et al. 2011.

2 The AiF is the umbrella organisation of 101 independent industrial research associations that organize pre-competitive research, each operating at sectoral industry level.

Thus, both programmes can be classified as systemic innovation policy instruments that can be adequately evaluated by employing systems evaluation approaches.

6 WHAT IS "SYSTEMS EVALUATION"?

We define "system evaluations" as evaluation approaches that give central importance to the investigation of the systemic interrelationships that determine the development of the object of study. They aim to do justice to the complexity of the object under investigation and are thus characterized by a holistic perspective. Central features of systemic evaluations are the critical reflection of the perspectives adopted in the analysis of the object and the definition of the boundaries of the system to be examined (this aspect particularly emphasised in Hummelbrunner 2011; Williams 2015). Systems evaluations are characterized by a perspective on the whole of government intervention concerned, which places the expected or observed effects of the intervention in the larger context of the respective policy and, on this basis, examines their meaningfulness and relevance.

Consequently, a systems evaluation is not a specific evaluation method that can be placed alongside other methods such as RCTs or peer interviews. It is not defined by the application of this or that exclusive method or, in the case of multi-method designs, of a specific set of methods. In principle, systemic evaluations can make use of the entire arsenal of quantitative and qualitative methods commonly used in evaluation research. Nevertheless, not every combination of methods is equally suitable to support systemic evaluation.

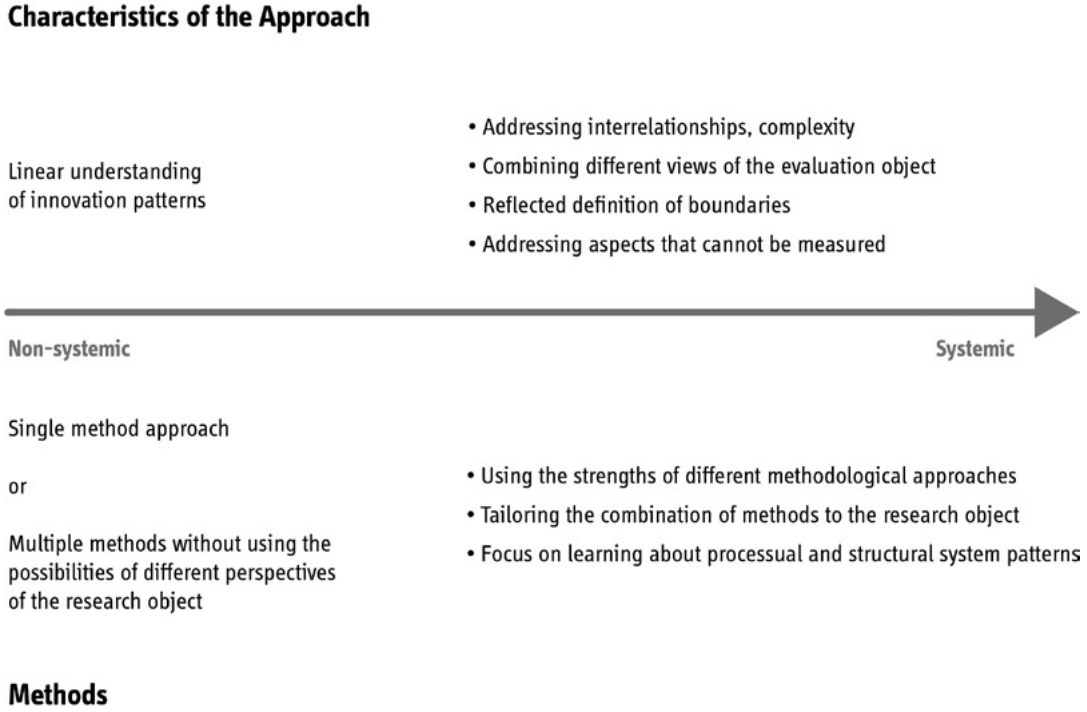
A systemic evaluation considers such development potentials that go beyond the framework of simple, linear causal relationships between the elements of the system. It is appropriate whenever the object to be evaluated is situated in a systemic context that is essential for understanding the effectiveness of the policy intervention and the mechanisms that it intentionally or unintentionally triggers.

The question arises, when an investigation takes on a "systemic character" or when does it definitely lose it? This question is relevant for all disciplines of scientific research that deal in any form with complexity. If the researcher opts for a rather generous answer, any investigation reveals systemic qualities. In order to bring a clear analytical line into the assessment of the systemic qualities of studies, one must be guided by plausible criteria that can be easily reproduced and applied by everyone. It is wrong to assume a binary classification, such that some studies do not meet the requirements of systemic analysis in any way, while others do so completely. Rather, we are dealing with a continuum of more or less strong systemic traits (Figure 3).

Decisive criteria for the inclusion of a systemic perspective in an evaluation design should be, in particular, the positive response to following questions:

- Is the realm of reality in which the intervention takes place characterised by properties that are typical of complex adaptive systems such as complexity, non-linearity, self-organization and emergence?
- Are expectations regarding desired outcomes of the intervention characterised by a high degree of uncertainty?
- Are serious results of the intervention reasonably to be expected only after long periods of time and dependent on many external and internal factors of the system, including its self-organising processes, which cannot be controlled by the policy-maker?
- Can consideration of the different perspectives of the programme makers and stakeholders involved contribute significantly to a better understanding of the programme?
- Does the delineation of the boundaries of the object of evaluation raise questions that are relevant to the assessment of the intervention?

Figure 3: Characteristics of systemic and non-systemic evaluations approaches
Source: Own depiction.



7 CAUSES OF LOW RECEPTIVENESS TO SYSTEMS THINKING

As we discussed above, elements of systems thinking can certainly be found in the practical evaluation of innovation policy measures, but there is hardly any question of a broad reception and general use of corresponding research instruments in projects predestined for this purpose. Barriers to a stronger echo of systems and complexity research in the evaluation of innovation policy measures are to be found in both epistemological and institutional fields. This is a topic that would deserve a substantially more elaborate discussion than is possible here. Some probably important aspects are elaborated here.

The human mind is primarily calibrated to the perception of simple, linear causal chains and is inclined to reproduce the mental model of the perception of linear causalities that proves itself anew every day. This serves as a ubiquitous model of knowledge even where it is not or only partially suitable for understanding a situation and often leads to erroneous attribution of blame for developments that have complex causes. Moreover, the evolution of the development of the natural sciences in western societies since the scientific revolution has fostered a type of technical rationality that has reinforced the predominance of linear causal thinking to the detriment of a willingness to adopt holistic perspectives (Meadows 2008). Of course, systems and complexity theory itself is a result of the development of technical rationality and arose in the mid-20th century in an effort to solve complex practical problems using sophisticated mathematical methods.

It should be noted, however, that systemic thinking, although by no means completely alien to people, tends to lose out in everyday life in comparison to linear thought patterns for epistemological and cultural reasons (Beasley 2012; Dörner 1997; Meadows 2008: 4), which is equally true for lay and professional people. The observation that the approaches of most innovation policy measures in the industrialised countries still follow the linear model of the innovation process today, despite its abandonment by innovation research decades ago, is probably largely due to this. Against this background, the spread of systems thinking in evaluators' community, but also in state bureaucracies, and above all its active use in practical work, does not happen automatically.

Another epistemological factor lies in the nature of the basic discipline of systems analysis and the particularities of its application. Systems and complexity research uses complicated mathematical models and is therefore easily suspected of being a playground for the mathematically gifted, who are likely to be found among practical programme evaluators only to a limited extent. There is a fundamental misunderstanding here. There is obviously no necessity to make system evaluations of government interventions dependent on the development of sophisticated mathematical models of the object of evaluation. In most cases, such an attempt would not be justifiable in any way in terms of financial and human resources. Rather, as Arnold (2004) shows, the use of system dimensions can be designed very differently depending on the object of study and the evaluation context. Incidentally, Bonini's paradox applies[3]: The more one tries to reflect complex systems in mathematical models by capturing as many relevant elements and relationships as possible, the less suitable this increasingly realistic model is for understanding reality.

Another factor that may be important in explaining the low reception of the systems concept are the worldviews sometimes conveyed together with it. As justified as the call for taking into account different perspectives on the object of evaluation and the emphasis on the role of boundary setting are (e.g., Williams 2015), they easily expose themselves to suspicion of ideology in practical contexts. The fact that the recent unprecedented rise of systems thinking is closely linked to the climate policy debate (e.g., Ison and Straw 2020; Meadows 2008) does not necessarily make things easier.

In addition to epistemological barriers, institutional barriers must also be taken into account. The mostly hard departmental demarcation between the ministries and the internal structuring of tasks in the ministries into clearly defined areas does not necessarily make it easier to incorporate concepts that imply cross-ministerial and cross-departmental cooperation. Proposals from outsiders who want to shake up the established boundaries of the areas of responsibility are not necessarily received with enthusiasm. In this respect, political decision-makers' declarations of intent to take a more systemic approach in the future should be treated with caution.

State bureaucracies have an apparent ability to take unwelcome ideas on the surface and distort them beyond recognition in the political process. An example of this is the 50-year "struggle" of the different German federal governments against bureaucracy, documented in a multitude of political documents, which continues to flourish and thrive despite all efforts in this regard. The same could happen with systemic thinking. This argument is not about finding culprits - this would be an expression of linear causal thinking -, but about elementary mechanisms of the development of bureaucracies.

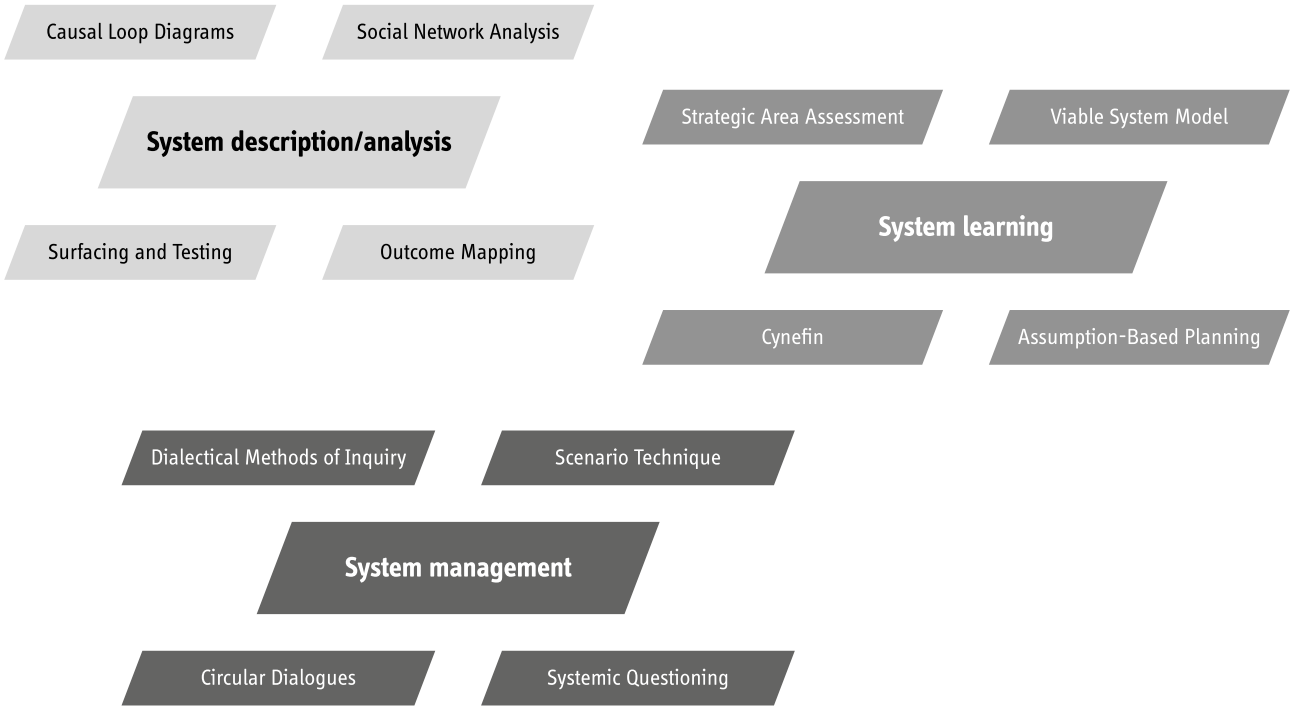
8 SUMMARY AND PROPOSALS FOR A PRAGMATIC USE OF SYSTEMS APPROACHES

Although the receptiveness for the inclusion of systemic perspectives in research has increased strongly in many scientific fields in recent decades, the potentials inherent in such approaches have so far only been used to a rudimentary extent in the practical evaluation of innovation policy programmes. This applies equally to assessments of innovation policy as a whole, the application of systemic perspectives to the simultaneous use of different, complementary policy instruments to achieve complex objectives, and to the evaluation of individual innovation policy programmes, such as the promotion of technology clusters, which target a complex object in order to pursue complex economic and technological

3 See wikipedia: „Bonini's paradox. Retrieved from https://en.wikipedia.org index.php?title=Bonini%s_paradox&oldid=1029627549, 21 June 2021.

objectives. Deficits in the first two mentioned areas have been repeatedly pointed out recently (e.g., Borrás and Laatsvit 2019; Edler and Fagerberg 2017). Our contribution aims to draw attention to the inclusion of systemic perspectives (a system-oriented framing) to the evaluation of individual programmes. This systemic framing can make useful contributions, and in many cases, it is only from this that a deep understanding of the state interventions in question can be generated.

Figure 4: The methodological arsenal of system-oriented evaluations



Cynefin is a knowledge management model developed by David Snowden and Cynthia Kurz for the analysis of complex adaptive systems. It is based on a typology of situations (simple, complicated, complex, chaotic), which takes a mediating position between complexity-reducing and complexity-emphasising procedures. The model takes into account the uncertainties inherent in complex adaptive systems that arise in the analysis and decision-making process (Williams and Hummelbrunner 2011: 163-183).

Source: Own depiction based on Williams and Hummelbrunner 2011.

Systemic evaluation amounts to a full consideration of the complexity properties of the object of study. It will always be part of a comprehensive methodological design that provides for the triangulation of the methodological tools contained in the mixed-method design. Therefore, under the umbrella of a systemic evaluation, rigorous quantitative methods such as RCTs will be found as well as qualitative components. What should count is solely the suitability of the chosen combination of methods for the best possible fulfilment of the research task at hand. Depending on the research task and the research context, the specific methods that have been treated in the evaluation literature as an expression of systemic research approaches should also be taken into account (Figure 4).

A systemic perspective is in general compatible with other prominent evaluation approaches, not least with realist evaluation and the diverse approaches of theory-based evaluation (Giel 2013, on the combination with complexity theory-based approaches Stame 2004). Both concepts have provided essential impulses for practical evaluation in multiple policy areas. Ray Pawson, pioneer of realist evaluation, seems to be sceptical of competing evaluation approaches that claim to do justice to the complexity of the evaluation object, including the systemic perspective (Pawson 2013: 53ff.). Other authors, however, point to the compatibility of systems/complexity approaches with realist evaluation (Westhorp 2012).

There are some practical steps that would lead practical evaluations to come closer to the idea of systemic evaluation:

i) Tailoring an evaluation programme that fits the characteristics of the object of evaluation in the best possible way;
ii) using instruments of complexity research in a complementary and supplementary function in normal routine evaluations of systemic evaluation objects;
iii) building experience driven models of the object of investigation without the use of overly complex-theoretical and mathematical constructions; this also means taking into account the complexity of interdependencies, uncertainties and emergent processes that lead to results of innovation funding;
iv) combining a system-oriented framing of an evaluation with all conventional (quantitative and qualitative) evaluation methods.

A systems approach can prove useful even in the case of rather simple innovation programmes (like "go-cluster").

Although for logical-systematic reasons there can be no evaluation that does not address systemic aspects in some way, the comprehensive consideration of dynamic systemic relationships in innovation policy evaluations has so far remained an exception. As we have tried to show, both epistemological and institutional factors are responsible for the unwillingness to adopt a systemic perspective in innovation policy evaluations.

It seems likely that evaluators of complex innovation programmes will claim that they were aware of the complexity of their object of study in the work process. They furthermore would make practical efforts to do justice to this complexity in the construction of their method design as well as in the practical evaluation work, at least insofar as the practical circumstances permit this. The lack of access to relevant data, the limited resources available and the time constraints of the evaluation alone would not have allowed this. An uneasy feeling might remain that essential things about the object of evaluation have not been revealed in the resulting research reports. Arnold (2004) has demonstrated that systemic perspectives can be introduced into such analyses even under conditions of numerous restrictions.

REFERENCES

Ackoff, R.L. (1971) Towards a System of Systems Concepts. Management Science 17(11): 661-671.

AEA (2015) Evaluating Outcomes of Publicly-Funded Research, Technology and Development Programs: Recommendations for Improving Current Practice. Version 1.0. Prepared by the Research, Technology and Development Evaluation Topical Interest Group of the American Evaluation Association (AEA), February 2015. Download under www.eval.org 21 January 2020.

Arnold, E. (2004) Evaluating research and innovation policy: a systems world needs systems evaluations. Research Evaluation (13)1: 3-17. Download from https://academic.oup.com/rev/article-abstract/13/1/3/1593255, on 24 February 2020.

Barbrook-Johnson, P., B. Castellani, D. Hills, A. Penn and N. Gilbert (2021) Policy evaluation for a complex world: Practical methods and reflections from the UK Centre for the Evaluation of Complexity across the Nexus. Evaluation 27(1): 4-17.

Beasley, R. (2012) 4.3.1 The Barriers to Systems Thinking. INCOSE International Symposium – July 2012. DOI: 10-1002/j.2334-5837.2012.tb01353.x.

Borrás, S. and J. Edler (2020), The roles of the state in the governance of socio-technical systems' transformation. Research Policy 49(5) 103971: 1-9.

Borrás, S. and C. Edquist (2019) Holistic Innovation Policy. Theoretical Foundations, Policy Problems, and Instrumental Choices. Oxford: Oxford University Press.

Borrás, S. and M. Laatsit (2019) Towards system oriented innovation policy evaluation? Evidence from EU28 member states. Research Policy 48(1): 312-321.

Bush, V. (1960) [1945]. Science – the endless frontier. A report to the President on a Program for Postwar Scientific Research. July 1945. Reprinted July 1960. Washington, D.C.: National Science Foundation.

Chaminade, C., B-A. Lundvall and S. Heneef (2018) Advanced Introduction to National Innovation Systems. Elgar Advanced Introductions. Cheltenham, UK, Northampton, USA: Edward Elgar.

Deutscher Bundestag, 19. Wahlperiode (2019) Bericht über die Programme zur Innovations- und Technologieförderung im Mittelstand in der laufenden Legislaturperiode, insbesondere über die Entwicklung des Zentralen Innovationsprogramms Mittelstand (ZIM). Fortschrittsbericht für das Jahr 2018. Unterrichtung durch die Bundesregierung. Drucksache 19/14480. Berlin: Deutscher Bundestag.

Dörner, D. (1997) [1089] Die Logik des Mißlingens. Strategisches Denken in komplexen Situationen. Rororo science. Reprint of paperback edition. Reinbek bei Hamburg: Rowohlt. Taschenbuch Verlag.

Easterly, W. (2007) [2006] The White Man's Burden. Why the west's efforts to aid the rest have done so much ill and so little good. Paperback edition. Oxford: University Press.

Edler, J., J. Fagerberg (2017) Innovation policy: what, why, and how. Oxford Review of Economic Policy. 33(1): 2-23. Download from: https://academic.oup.com/oxrep/article-abstract/33/1/2/2972712, on 19 November 2019.

Edquist, C. (2005) Systems of Innovation: Perspectives and Challenges. In J. Fagerberg, D.C. Mowey and R.R. Nelson (Eds.) The Oxford Handbook of Innovation. Oxford: Oxford University Press, pp. 181-208.

Eckert, S., V. Schüren and A. Bode (2016) Evaluation des Programms go-cluster des Bundesministeriums für Wirtschaft und Energie (BMWi). Abschlußbericht Berlin und Bickenbach: INTERVAL GmbH and CONABO UG (haftungsbeschränkt).

Forss, K., M. Marra and R. Schwartz (Eds.) (2017) [2011] Evaluating the Complex. Attribution, Contribution, and Beyond. Comparative Policy Evaluation 18. London and New York. Routledge.

Forss, K. and R. Schwartz (2017) [2011] Introduction. In Forss, K., M. Marra and R. Schwartz (Eds.) Evaluating the Complex. Attribution, Contribution, and Beyond. Comparative Policy Evaluation 18. London and New York. Routledge, pp. 1-32.

Giel, S (2013) Theoriebasierte Evaluation. Konzepte und methodische Untersuchungen. Internationale Hochschulschriften 584. Münster: Waxmann.

HM Treasury (2020) Magenta Book 2020. Supplementary Guide: Handling Complexity in Policy Evaluation. London: HM Treasury.

Holland. J.H. (2014) Complexity. A Very Short Introduction. Oxford: Oxford University Press.

Hummelbrunner, R. (2011) Systems thinking and evaluation. Evaluation 17(4): 395-403, DOI:10.1177/1356389011421935.

Imam, I., A. LaGoy and B. Williams (2006) Introduction. In Williams, B. and I. Imam (eds.) (2006) Systems Concepts in Evaluation: An Expert Anthology. American Evaluation Organisation, pp. 3-10.

Ison, R. and E. Straw (2020) The Hidden Power of System Thinking. Governance in a Climate Emergency. London and New York: Routledge.
Lundvall, B.A. (2010) National Systems of Innovation. Toward a Theory of Innovation and Interactive Learning. The Anthem Other Canon Series. London, New York, Delhi: Anthem Press.

Mainzer, K. (2008) Komplexität. UTB 3021. Paderborn: Wilhelm Fink.

Mainzer, K. (2014) Die Berechnung der Welt. Von der Weltformel zu Big Data. München: C.H.Beck.

Meadows, D.H: (2008) Thinking in Systems. A Primer. Edited by Diana Wright, Sustainability Institute. White River Junction, Vermont: Chelsea Green Publishing.

Midgeley; G. (2006) Systems Thinkling for Evaluation. In: B. Williams and I. Imam (Eds.) Systems Concepts in Evaluation: An Expert Anthology. American Evaluation Organisation

Mitchell, M. (2009) Complexity. A Guided Tour. Oxford, New York: Oxford University Press.

Mowles, C. (2014) Complex, but not quite complex enough: The turn to the complexity sciences in evaluation scholarship. Evaluation 20(2): 160-175.

Moyo, D. (2009) Dead Aid. Why aid is not working and how there is another way for Africa. London: Penguin.

Pawson, R (2013) The Science of Evaluation. A Realist Manifesto. Los Angeles et al.: Sage.

Rothgang, M., M. Peistrup and B. Lageman (2011) Industrial Collective Research Networks in Germany: Structure, Firm Involvement and use of Results. Industry and Innovation 18(4) 393-414.
RWI and WSF (2010) Erweiterte Erfolgskontrolle beim Programm zur Förderung der IGF im Zeitraum 2005 bis 2009. Endbericht. Forschungsvorhaben des Bundesministeriums für Wirtschaft und Technologie. Essen: RWI und WSF.

Schot, J. and Steinmueller, W.E. (2018). Three frames for innovation policy. R&D, systems of innovation and transformative change. Research Policy 47: pp. 1554-1567.Smits, R. and S. Kuhlmann (2004) The rise of systemic instruments in innovation policy. International Journal of Foresight and Innovation Policy. 1(1/2): 4-32.

Stame, N (2004) Theory-based Evaluation and Types of Complexity. Evaluation 10(1): 58-76.

Uyarra, E. and R. Ramlogan (2017) Cluster policy in an evolutionary world? Rationales, instruments and policy learning. In D. Fornahl and R. Hassink (Eds.) The Life Cycle of Clusters. A Policy Perspective. Cheltenham, UK, Northampton, MA, pp. 35-55: Edward Elgar.

Williams, B. (2015) Prosaic or Profound? The Adoption of Systemic Ideas by Impact Evaluation. IDS Bulletin 46(1): 7-16.

Williams, B. and R. Hummelbrunner (2011) Systems Concepts in Action: A Practitioner's Toolkit. Stanford, CA: Stanford University Press.

Williams, B. and I. Imam (eds.) (2006) Systems Concepts in Evaluation: An Expert Anthology. American Evaluation Organisation.

AUTHORS

MICHAEL ROTHGANG
RWI - Leibniz Institute for Economic Research
Hohenzollernstr. 1-3, 45128 Essen, Germany
E: michael.rothgang@rwi-essen.de
ORCID ID: 0000-0001-8328-0295

BERNHARD LAGEMAN
RWI - Leibniz Institute for Economic Research
Hohenzollernstr. 1-3, 45128 Essen, Germany
E: bernhard.lageman@t-online.de

KEYWORDS: Systems evaluation; innovation policy; systems analysis; complexity; innovation policy evaluation

WORKING ON IMPACT AND CONTRIBUTING TO RESEARCH & INNOVATION POLICIES – LOOKING BACK AND AHEAD

BETTINA UHRIG AND BARBARA SPANÓ

DOI: 10.22163/fteval.2022.557

ABSTRACT

This article presents impact case studies at research project and organisational levels by exploiting the Horizon Europe concept of pathways to impact and the proposed indicators. In Horizon Europe, which is the European Commission's funding programme for research and innovation, time-sensitive Key Impact Pathways and related indicators are used as a tool for assessing the different types of impact: scientific, societal, and economic.

Based on many years of experience with stakeholder engagement and impact, the authors focus on the indicators for assessing societal impact. In this way, the authors would like to contribute to the discussion on creating societal impact through research projects and institutional strategies. Leading questions are 1) Can Research & Innovation (R&I) policies be improved by using Horizon Europe Key Impact Pathways and related indicators? And 2) Can an institutional impact project and even a research project benefit from using Horizon Europe indicators and at the same time feed into R&I policies?

WORKING ON IMPACT AND CONTRIBUTING TO R&I POLICIES – LOOKING BACK AND AHEAD

1. OBJECT AND PURPOSE

The main focus of this paper is on the new Key Impact Pathways (KIPs) and their related indicators used in Horizon Europe (HE[1], see: European Commission, 2018a). The purpose is to explore whether the application of KIPs and their indicators can encourage the discussions on societal impact and can be used to assess pathways towards societal impact created by a research project (first case study) as well as an institutional impact project (second case study) and, by doing this, feed into the discussions on R&I policies.

The HE Framework Programme (FP) has a stronger focus on impact than its predecessor, Horizon 2020 (H2020). By investing in areas that are of key strategic interest for Europe, HE frames and stresses how the impact of research and innovation can contribute to the implementation of the policy priorities of the European Union (EU) as well as to the achievement of the United Nation's Sustainable Development Goals.

The impact approach of HE

> *"aligns with a new level of ambition to boost the diversity of impact of EU research and innovation funding. The objective is to allow policy makers and the wider public to get regular insights regarding the effects and benefits of the programme or European science, the economy and wider society."*[2]

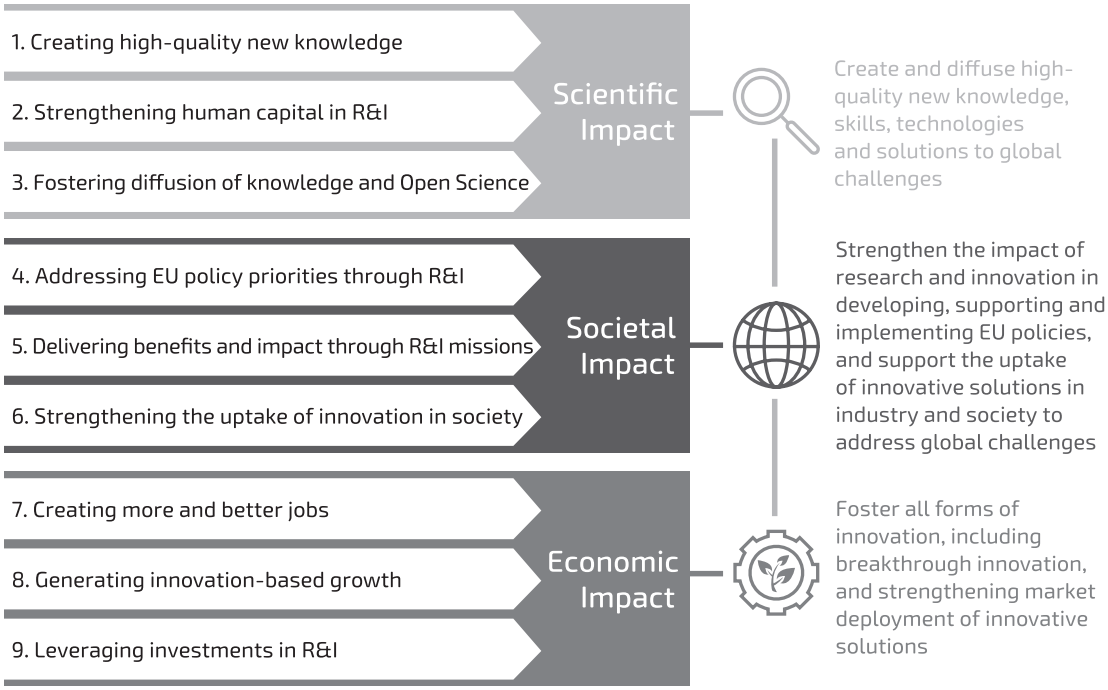
To monitor this approach, the European Commission (EC) has agreed on new KIPs, a concept that will also be used by the authors:

> *"The HEU legislation includes an obligation to monitor the effectiveness of measures to improve citizen and civil society involvement. This is where the new Key Impact Pathways (KIPs) come in: In the HEU Impact Assessment, the EC identified nine KIPs for the future FP, which are subsumed in three categories – scientific, societal, and economic impacts. KIPs will replace the Horizon 2020 Key Performance Indicators (KPI). KIPs and related KIP indicators will structure the monitoring of the FP's progress towards its objectives. The KIPs, so the EC, stem from a need to better communicate this progress and to better demonstrate why EU R&I investments matter. Representing the 'backbone of the HEU monitoring and evaluation', the corresponding KIP indicators will unite both qualitative and quantitative information and will be reported on an annual basis." (SwissCore 2020, p.18)*

1 Two acronyms are used for Horizon Europe: HEU or HE. The EC mostly uses HE, which will be the acronym used in this paper, except in quotations.

2 European Commission (2021), webpage: Horizon Europe programme analysis, viewed February 16, 2022, https://ec.europa.eu/info/research-and-innovation/strategy/support-policy-making/shaping-eu-research-and-innovation-policy/evaluation-impact-assessment-and-monitoring/horizon-europe_en.

Figure 1: Three types of impact, tracked with KIPs (European Commission 2018a, p.104)



The KIPs set the frame for detailed indicators, which have been developed to address the specificity of the different actions that constitute the programme. The KIPs are time-sensitive, and the time aspects of 'societal impact pathways indicators' are shown in Figure 2:

Figure 2: Societal impact pathway indicators (European Commission, 2018b, page 16)

Towards societal impact	Short-term	Medium-term	Longer-term
Addressing EU policy priorities through R&I	Outputs - Number and share of outputs aimed at addressing specific EU policy priorities	Solutions - Number and share of innovations and scientific results addressing specific EU policy priorities	Benefits - Aggregated estimated effects from use of FP-funded results, on tackling specific EU policy priorities, including contribution to the policy and law-making cycle
Delivering benefits and impact through R&I missions	R&I mission outputs - Outputs in specific R&I missions	R&I mission results - Results in specific R&I missions	R&I mission targets met - Targets achieved in specific R&I missions
Strengthening the uptake of innovation in society	Co-creation - Number and share of FP projects where EU citizens and end-users contribute to the co-creation of R&I content	Engagement - Number and share of FP beneficiary entities with citizen and end-users engagement mechanisms after FP project	Societal R&I uptake Uptake and outreach of FP co-created scientific results and innovative solutions

To conceptualise and contribute to the discussion of societal impacts generated at the regional, national and EU levels, the authors present two case studies reflecting on the achievements of these projects and the time-sensitive KIP indicators. 'Short-term' captures the time during the implementation of the project and up to three years after the end of the project, 'medium-term' the period of three to five years after the end of the project and 'longer-term' relates to societal impact achieved more than five years after the end of the project.

2. CASE STUDIES OF WORKING WITH IMPACT

Policy makers, public research-funding bodies, like the EC and the Research Council of Norway, and private research-funding bodies, like the Danish independent foundation Novo Nordisk Foundation, have increased their focus and set their expectations for research institutions to demonstrate the immediate and long-term societal and economic impact of research. This has required, and still requires, a change in the way researchers, research managers and universities think about and understand the effects of research and how research is performed if they want to be competitive not only in securing research funding, but also in attracting students and staff.

The two case studies represent two very different ways of working on impact:

1. for the H2020 project, it was a prerequisite to work with measures which can lead to societal impact, and monitoring the possible impact has been one of the tasks for the project's Impact Manager (Bettina Uhrig);
2. for the impact project at the university, one of the main objectives was to empower researchers to enhance the benefits of research for society.[3]

Both case studies are written by the authors of this article, the H2020 project's Impact Manager and the initiator and adviser for the impact project at the university. The case studies are reports of measures which can contribute to achieving societal impact. The authors wrote these reports based on their experiences and tasks related to each project. The case studies illustrate the practical work with supporting pathways to impact. Pathways to impact are a concept used in HE and are defined as

> *"logical steps towards the achievement of the expected impacts of the project over time, beyond the duration of a project. A pathway begins with the projects' results, to their dissemination, exploitation and communication, contributing to the expected outcomes in the work programme topic, and ultimately to the wider scientific, economic and societal impacts of the work programme destination". (European Commission 2021c, page 29).*

HE started in 2021 and evaluations of the programme are not yet available, therefore the authors cannot refer to such evaluations, instead they would like to contribute to a discussion of HE and its impact pathways. In doing so, the authors hope to contribute to the wider debate on understanding the impact of research relevant for policy (see: Williams and Lewis 2021).

2.1 THE H2020 PROJECT CASE STUDY: NATIONAL STAKEHOLDER GROUPS IN DARE[4]

The project Dialogue about Radicalisation and Equality (DARE)

> *"aimed to deepen our understanding of radicalisation through a critical and societally focused approach. Funded under the EU Horizon 2020 Framework Programme for Research and Innovation, DARE investigated young people's encounters with radical(ising) messages, how they responded to such calls, and the choices they made about the paths they took. The project undertook extensive empirical research with young people in radical(ising) milieus both offline and online and generated important insights into what drives radicalisation but also what constrains it. The findings suggest that the situated knowledge of actors in radical(ising) milieus might be utilised to prevent and counter extremism"*[5]

The call topic behind the project with the title 'Contemporary radicalisation trends and their implications for Europe' was part of the H2020 Work Programme 2016-2017 for Societal Challenge 6 'Europe in a changing world – Inclusive, innovative and reflecting societies' and belonged to the call 'Reversing inequalities'. In the topic description it was stated that *"radicalisation is on the rise"* and that *"research under this topic will considerably enhance the knowledge base on the scope, origins, causes and cognitive as well as emotional dynamics of radicalisation"* and will through its results impact on future policies preventing radicalisation, which was described as expected impact (European Commission 2017, page 36).

DARE, the only project funded under this topic, was implemented from May 2017 until October 2021, and coordinated by the University of Manchester. DARE comprised 17 organisations from 13 different countries[6]. To secure the collaboration with stakeholders, which is seen as an important tool for pathways towards impact, the DARE consortium had agreed on establishing National Stakeholder Groups (Uhrig 2019/2020).

3 The other objective was to remain competitive.

4 This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 725349. The case study is based on discussions with DARE colleagues (see also Figure 4), internal impact reports and a public Deliverable written by the Impact Manager (Bettina Uhrig) for the DARE project. The Impact Manager was supported by the Impact Sub-Committee consisting of her, the Project Manager and two to three researchers involved in DARE.

5 DARE, Introduction to the programme of the Research-Policy-Practice Event, 21 -23 September 2021, online, website no longer available.

6 http://www.dare-h2020.org/
Consortium Members: The University of Manchester, United Kingdom (Coordinator); Oslo Metropolitan University, Norway; École des Hautes Études en Sciences Sociales, France; Anadolu University, Turkey; German Institute for Radicalization and Deradicalization Studies, Germany; Leiden University, The Netherlands; Hochschule Düsseldorf – University of Applied Sciences, Germany; Teesside University, United Kingdom; Collegium Civitas University, Poland; Panteion University of Social and Political Science of Athens, Greece; Higher School of Economics, Russia; The Institute of Social Sciences Ivo Pilar, Croatia; European Network Against Racism, Belgium; The People for Change Foundation, Malta; Sfax University, Tunisia; University of Oslo, Norway; University of Birmingham, United Kingdom.

NATIONAL STAKEHOLDER GROUPS – DESCRIPTION

The DARE consortium had agreed that National Stakeholder Groups (NSGs), consisting of a broad range of relevant policy-practitioner and scientific partners, should

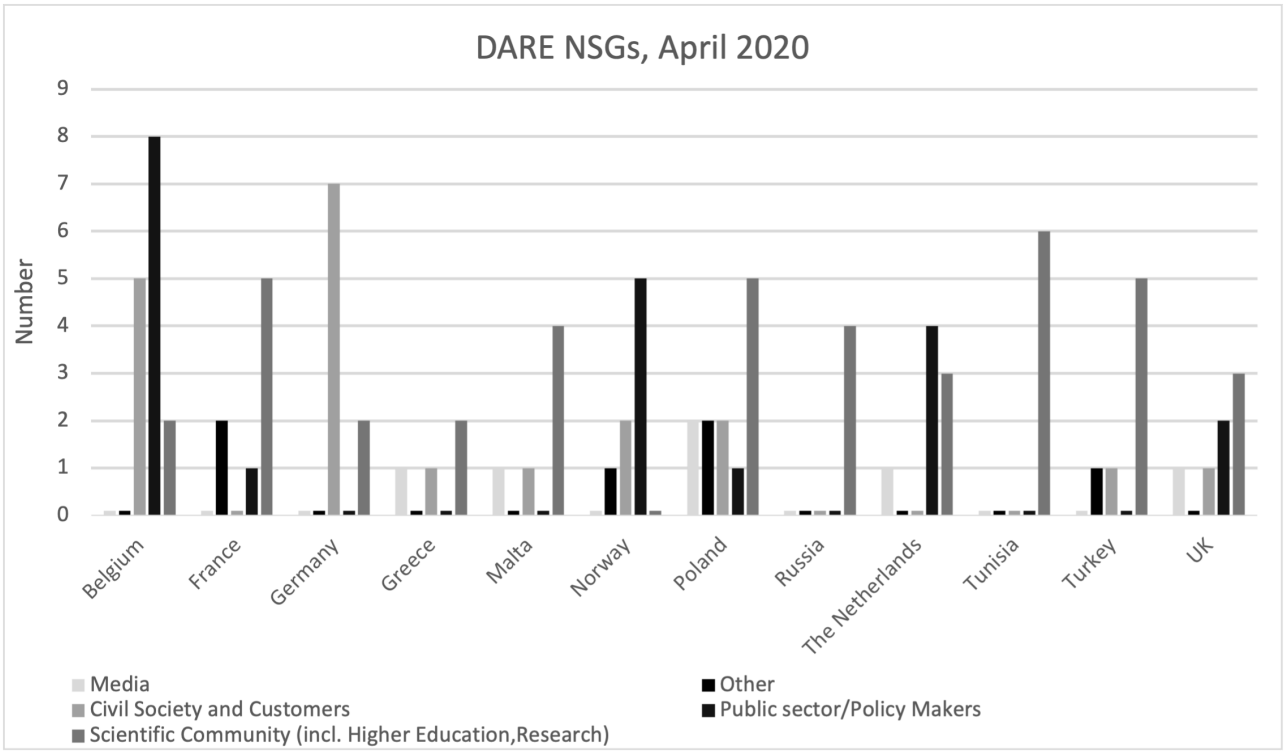
"meet regularly to:

i) advise on the development of the research,
ii) discuss the significance of emerging findings,
iii) advise on the production of Policy Briefs and Recommendations, and
iv) facilitate the dissemination of research findings into policy arenas at local, national and European levels."[7]

The NSGs were foreseen in all DARE countries, except Croatia, where no fieldwork was planned. All DARE partners responsible for a NSG, wrote minutes of their NSG meetings, which were collected by the DARE Impact Manager, who was monitoring the work with the NSGs. By April 2020, NSGs had met at least once in all 12 countries; altogether, 27 NSG meetings took place between May 2017 and April 2020. With the start of the pandemic, physical meetings were no longer possible. However, between May 2020 and October 2021 six virtual NSG meetings and one face-to-face meeting took place in four different countries (Germany, Norway, Poland, and the United Kingdom). Furthermore, DARE colleagues had individual online meetings with NSG members to discuss the progress of DARE as well as dissemination activities at national and European level. For example, several NSG members from different countries (France, Germany, Norway, and the United Kingdom) were involved in presenting their work at the virtual DARE Research-Policy-Practice Event from 21st to 23rd September 2021.

The NSGs varied in the types of stakeholders involved. In countries such as Turkey and Tunisia where radicalisation is a highly political and contentious topic, NSGs were comprised mainly of academic members. In Belgium, Norway and the Netherlands, employees from the public sector and policy makers were the dominant group. Over the course of the project, the composition of the NSGs changed, some members left because they retired or moved to a new job (or because the employer did not agree to their participation in an NSG related to radicalisation) and new members with a different background joined the NSGs. The NSGs, which started in 2017 and 2018, had fewer active members in 2020 and 2021. In spring 2020, most NSGs had between four to eight members. All these changes influenced the discussions of possible dissemination and exploitation actions leading to impact. The total number of members was 94 (not counting the DARE team members), nearly as many women as men were members in the NSGs (50 male and 44 female). The following figure illustrates the diversity of NSG members in the different DARE countries.

Figure 3: Diversity and number of the NSG members in the different countries[8]



7 DARE (2020), Description of Action, page 107. Not public.

8 The classification follows the EC reporting system for H2020 projects. Representatives from industry and investors were not members in any of the NSGs. The figure shows the diversity and numbers of the NSG members based on reports from NSG meetings from the start of the project (May 2017) until the beginning of the pandemic (April 2020).

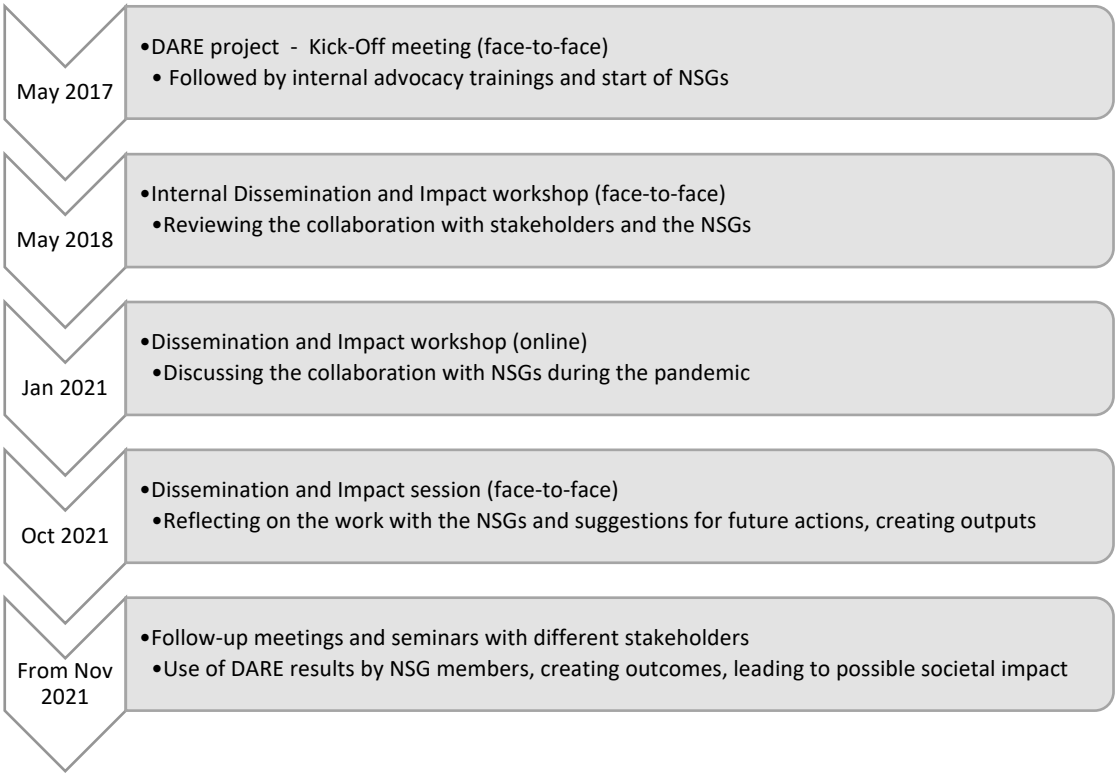
NSGS – RESULTS AND RECOMMENDATIONS

Establishing the NSGs and involving their members in discussions about DARE created inspiring dialogues with stakeholders, even during the pandemic at online meetings.

The NSGs and their members contributed to the three main objectives related to DARE's collaboration with stakeholders: information-gathering, advisory and dissemination activities. NSG members, for example, supported researchers to find interviewees, participated in DARE events and contributed to the dissemination of DARE results.[9]

The following figure illustrates the development of the work with the NSGs:

Figure 4: Process of working with NSGs



During the last project meeting in October 2021, DARE partners concluded that NSGs which had mainly practitioners and/or academics as members were more stable: these stakeholders were interested in the project development and glad to contribute to it. For example, practitioners in Norway underlined that they appreciated the possibility to meet with other practitioners and academics and to have open discussions about their work. Involving policy makers as NSG members was experienced as 'difficult': they expect easy-to-communicate results, which are often only available at the end of the project.

As a H2020 project, DARE has not used the three HE KIP indicators, which were published after the start of DARE. However, relating DARE to the KIPs and the related indicators it can be stated that DARE has been working with some of the KIPs described for scientific and societal impact (see Figure 1):

> For example, DARE has created *new knowledge*, which has been used for writing peer-reviewed articles, research briefings, policy briefs and for producing films and educational toolkits.

Creating these outputs fits into the short-term societal impact pathway indicators (see Figure 2):

> Through its outputs DARE has addressed *EU policies* and involved *end-users*, e.g., as members of NSGs.

However, tracking impact after the end of the project is limited. We don't know yet how often DARE outputs will be used for creating outcomes10, for example for changing a de-radicalisation programme. It is even more difficult to track if DARE outputs and outcomes will lead to societal impacts, for example, if the toolkits will lead to combating radicalisation through dialogue. However, based on their experiences, the majority of DARE partners recommend NSGs as a tool for supporting pathways towards societal impact. Such pathways need to be described in HE proposals for Research and Innovation Actions (RIAs) and Innovation Actions (IAs) and must relate to the outcomes specified in the topic

9 These findings are based on the reports from NSG meetings and the evaluation of questionnaires sent to NSG members.

10 European Commission (2021c), page 29: Definition of outcome:
"The expected effects, over the medium term, of projects supported under a given topic. The results of a project should contribute to these outcomes, fostered by the dissemination and exploitation measures. This may include the uptake, diffusion, deployment, and/or use of the project's results by direct target groups. Outcomes generally occur during or shortly after the end of the project."

and the wider impacts specified in the respective destination of the work programme.

2.2 THE INSTITUTIONAL CASE STUDY: WORKING WITH IMPACT AT THE TECHNICAL UNIVERSITY OF DENMARK

Many universities nowadays face financial challenges, threatening both their ability to perform research and their role as part of the innovation ecosystem and global economy. Those challenges also affect their need to attract talented researchers, who can contribute with their knowledge and networks to develop a specific research area and to boost the universities' competitiveness. To address those challenges, universities have started implementing strategies and developing frameworks to increase their capability to attract external research funding. At the same time, funding bodies are focusing more on impact.

THE RESEARCH IMPACT PROJECT – DESCRIPTION

In May 2020, the Technical University of Denmark (DTU) launched a Research Impact Project as part of the university's broader sustainability programme[11]. The main programme objective was to promote a sustainable change in society through research, education, and innovation, and to create a more sustainable future. The project ran from May 2020 until May 2021 (see Figure 6) and is DTU's first attempt to address research impact at a corporate level.

The purpose of DTU's Research Impact Project was twofold: on the one hand, it aimed at providing the researchers with the necessary tools, guidelines and methodologies that could, inter alia, support them in addressing the societal and economic impact of their research to secure research funding and international competitiveness. On the other hand, it focused on empowering DTU's researchers in enhancing the benefits of their research outside academia. Figure 5 summarised the goal of the project and was used by the author to present it to the university leadership.

Figure 5: Purpose of the Research Impact Project

WHY?	HOW?
Empower the researchers to: • Enhance the benefits of their research on the society; • Remain competitive; • Secure funding.	• Understand impact and apply the principles in different contexts; • Start a discussion about the need of a definition; • Identify and develop tools for understanding and addressing impact.

Despite the fact that DTU's "Strategy 2020-2025 Technology for people" (Technical University of Denmark, n.d.) specifically mentions innovation and sustainability, DTU does not have an impact strategy or a dedicated research impact, communications and engagement team to ensure that the researchers' work has a broader reach and application that go beyond the research community. Therefore, to set a direction for the work on impact during the project, the project group agreed to use a working definition of impact that reflects DTU's strategy. The chosen definition is as follows:

> *"Impact is the provable effects of research in the real world; The changes we can see (demonstrate, measure, capture), beyond academia (in society, economy, environment) which happen because of our research (caused by, contributed to, attributable to); Driven by a number of factors including funders' requirements and research assessment." (See: Bayley and Phipps 2017, page 4)*

When the project was initiated, HE had not yet started. Despite knowing that there would be a paradigm change in the design of HE *"from an activity-driven to an impact-driven programme"* (European Commission 2021a, page 9) and despite being aware of the content of the "EUROPE Impact Assessment of the 9th EU Framework Programme for Research and Innovation" (European Commission 2018a), it was still not clear how the new impact design and the Key Impact Pathways would affect and would be critical for the content of the projects, the results, the expected outcomes and impacts. There was, however, sufficient impetus for many universities, and among them DTU, to start looking at impact in a more consistent way that could lead to addressing the targeted impact specified in the calls and topic texts. DTU's working definition was quite broad and did not take into consideration the time-sensitive aspects of impact and the differences between outcomes and impacts. These aspects emerged and became relevant during the project group's work, especially after obtaining more information on the development of the impact requirements in HE.

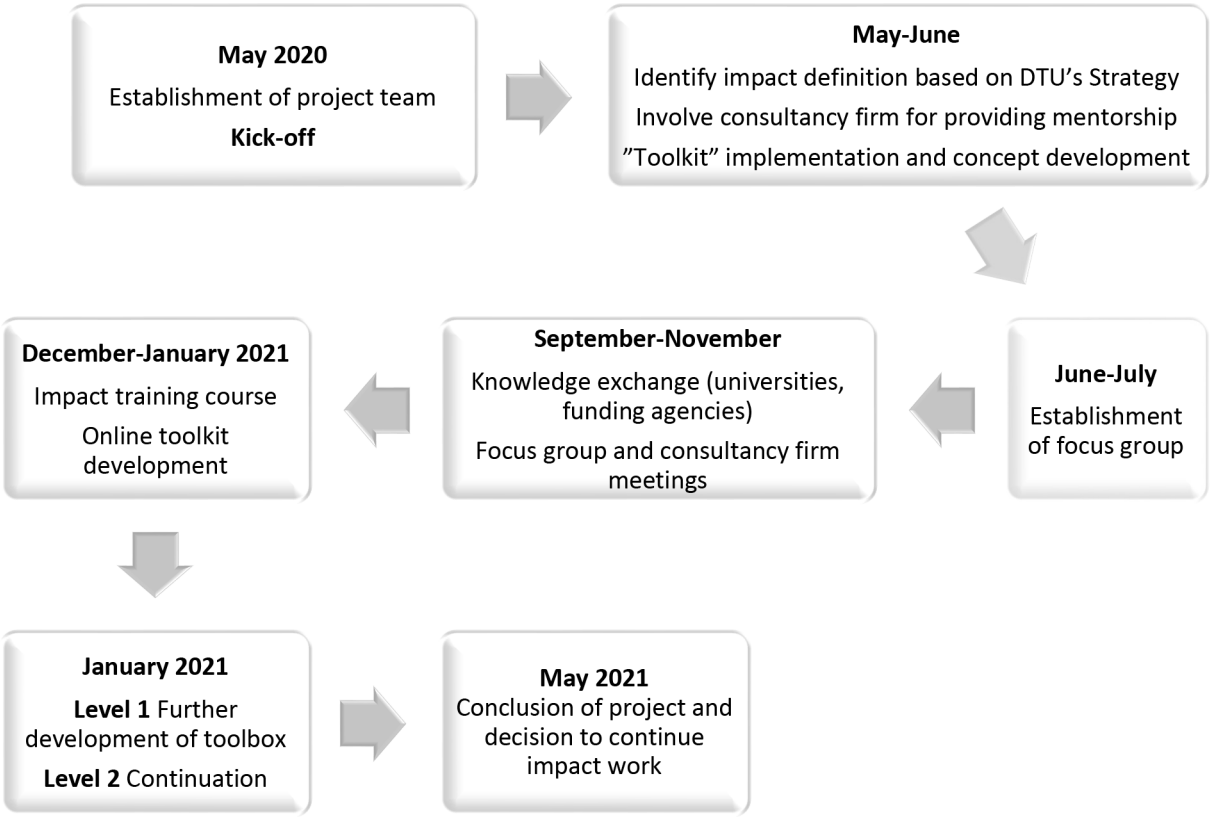
Having this in mind, the project group started to work in May 2020. Besides the co-author[12], it was composed of the Head of Office for Research, Advice and Innovation, DTU's Corporate Sustainability Manager, and two employees from the Office for Research, Advice and Innovation.

One group member focussed on the "elite/excellence grants" that traditionally reward researchers based on their scientific achievements and their scientific impact; the co-author and another member worked on the impact requirements for the "competitive applications". The two other members contributed with inputs to both areas and ensured that the work was streamlined with the entire sustainability programme.

11 The sustainability programme was an internal initiative, public documents are not available.

12 The co-author Barbara Spanó worked at DTU until October 2021.

Figure 6: Research Impact Project process



The work on impact in the "elite/excellence grants" one-person sub-group was closely coordinated from the start with the directors and the researchers of four designated departments dealing with areas like quantitative sustainability assessment, life cycle analysis and the circular economy.

On the other hand, the sub-group working on impact in the "competitive applications" started its work by involving a consultancy firm to provide a mentorship to tap into the members' existing knowledge, skills, and experience in working on stakeholder engagement and impact. The decision to have a mentorship programme was motivated by DTU's interest in ensuring that the competencies acquired could stay in-house and eventually lead to an impact team.

During the mentorship the discussion mainly focused on understanding the meaning of impact, impact pathways, how to address the funding agencies' requirements, the need to have an impact strategy, and how to support DTU's researchers and administration in addressing and communicating impact.

Based on those discussions, the sub-group decided to establish a small focus group comprising six researchers and research managers from the chosen departments. The focus group was tasked with identifying and describing the researchers' challenges when addressing impact in research applications. It was clear from the beginning that the group lacked understanding of what impact is about and how the requirements from the funding agencies frame the way impact is described in the proposal and is assessed by the evaluators. The time-sensitive aspects and, in this regard, the difference between outcomes and impact were also unclear. Confusion was also added because many funding agencies also use the word impact when referring to outcomes.

The project group members worked on average around 25 % of their time on the project, while the focus group members contributed around two hours a month. The project group and the focus group met around five times.

In addition to the focus group, many universities in Europe and Denmark as well as European and Danish private and public funding agencies contributed to the project by sharing with DTU their knowledge and approach in working with impact. They presented the mechanisms they established to support fundraisers and researchers to address impact and their requirements on how to address socio-economic impact in the medium and longer-term.

The feedback from the focus group and the learning from other institutions were taken into consideration in developing tools for supporting researchers and research managers in writing applications and will be taken into consideration if DTU decides to have a specific Impact Strategy[13].

13 In 2021, EARMA, the European Association of Research Managers and Administrators, published a study summarising results from a survey which aimed at identifying and sharing best practice in how institutions are responding strategically to the research impact agenda, and the ways in which this relates to national and European policy contexts (Jackson and Uhrig 2021).

THE RESEARCH IMPACT PROJECT – RESULTS AND RECOMMENDATIONS

The project was DTU's first attempt at tackling societal impact at the corporate level. It succeeded in raising awareness of societal impact both at the leadership level and with the researchers and fundraisers.

It provided support to the researchers, developed an understanding of research impact, collected and designed relevant tools, created a web page with impact-related documents, held and planned workshops and seminars.

After one year's work, the project's main findings can be summarised as follows:

- A *common understanding of impact at a fundamental level is needed*; however, the naturally associated processes and concepts are dynamic. Funding agencies and other stakeholders have specific and changing requirements regarding what can be considered "societal impact" as well as the way it should be described in the applications and reported both during and after the end of the projects. Universities need to constantly adjust their response to the specific requirements of the funding agencies by identifying relevant solutions and tools. How societal impacts can and should be achieved through research projects affects how results are communicated, disseminated, and exploited, which, in turn, changes how universities implement research and innovation.
- The *terminology of 'impact' and 'outcome'*: The differences between 'impact' and 'outcomes' in time, reach, scope and nature, are not always understood by researchers and are addressed in different ways by funding agencies.
- A *dedicated impact strategy*: Having an impact strategy shows a commitment from the leadership in handling research impact at the corporate level. Universities that have a specific impact strategy allocate more time and resources to supporting researchers in achieving societal impact. They have implemented structures and measures across departments (i.e., research, communication, partnerships, etc.) and disciplines, which ensures that the creation of societal impact is taken into consideration from the start, that relevant stakeholders are involved both during and after the end of the project, and that appropriate resources are provided. Training courses are offered to both academics and research managers to assess and boost their skills. There is a clear understanding of the roles and the level of support that can be expected at the institutional level.
- *Drivers*: Despite the interest the project received from researchers, fundraisers and department-level management, it can be difficult to activate and engage the researchers as many still perceive impact as something that lies on the periphery of their core tasks. Moreover, many funding agencies, public and private, focus only on research activities, making it harder for the researchers to understand the benefit of allocating time to activities like communication, dissemination and exploitation when applying for funding that has specific requirements in that direction. The project identified four main drivers: i) funders' requirements (e.g., impact as an evaluation criteria); ii) incentives at the university level (i.e., direct influence on the career of the researchers or financial benefit); iii) assessment requirements at the national level (e.g., the Research Excellence Framework, UK[14]); iv) the researchers' own clarity and commitment to social and economic change in the medium and longer-term.

Based on the above-mentioned findings, the project team decided to continue to work on impact at DTU. It assessed the need for constant capacity building among the researchers and the support staff as well as the need to involve the researchers themselves in the process. It also decided to eventually involve other expertise and offices like the bibliometric unit within the office for Research and counselling, the Library, and the office of Communication and Media tasked with developing and supporting DTU's external and internal communication. Finally, the project team agreed to define and develop indicators to measure the societal impact of research projects, starting with sustainability indicators.

3. CONCLUSIONS

Based on the way HE is conceived and its increased focus on impact, understanding the KIPs and the related indicators will be essential for the development and submission of excellent proposals and for the implementation of successful projects. Understanding the indicators and the way they will be used cannot only be a task for advisers and researchers working on HE proposals and projects. If HE and its projects are to create societal impacts at the regional, national and EU levels, the KIPs and the indicators as well as the policy behind them[15] have to be understood by the leadership of universities and other organisations involved in HE. Furthermore, the indicators will make it easier to compare possible societal impacts in different countries and to visualise the influence of HE and its projects on R&I and other EU policies.

The DTU case study focused on impact and KIPs at the proposal level and while doing that it started an institutional change process to better value the impact of research and innovation conducted at the university. The DARE case study focused on using NSGs as a tool for promoting impact pathways during the implementation of the project. In HE there is a clear expectation that the projects should continue to work on impact years after the end of the project itself. Measures like the Horizon Results Platform, the Horizon Impact Award, and the Innovation Radar offer support for working on impact and provide platforms where the results are shared and hopefully used by different stakeholders, aiming in this way at ensuring that impact can be achieved years after the end of a project.

This leads, though, to a series of open questions:

- How can the societal impact be tracked back to a specific research project? And is this the goal of the KIP indicators? Regarding the policy outcomes, it should also be noted that most high-level policy documents do not cite any sources and most of the impact in this sense is hidden in data sets like Overton[16].

14 https://www.ref.ac.uk/, viewed February 16, 2022.
15 See: European Commission (2021b), HORIZON EUROPE STRATEGIC PLAN 2021 – 2024, Brussels.
16 Overton website: Is your work influencing policy? https://www.overton.io/, viewed February 16, 2022.

- How will the EC support the beneficiaries in working on the medium-term and longer-term impact that happens after the end of the project?
- How can particular attention be given to exploiting results and information that can be used as an input to EU policymaking by the Commission Services and national administrations? This will necessitate better connections between implementing bodies and policymakers, including R&I supported under institutional partnerships.

Besides these questions, the case studies described above show that an institutional impact project and even a research project can benefit from using HE KIPs and their indicators and at the same time feed into R&I policies: the indicators can support the complex work with impact, can enrich the discussions, monitoring and reporting and can be useful for feedback to research-funding bodies.

4. REFERENCES

Bayley, Julie and Phipps, David (2017), Real Impact - Impact Literacy Workbook, Developed in conjunction with Emerald Publishing, no location.

European Commission (2021a), Horizon Europe (Horizon) Programme Guide, Brussels.

European Commission (2021b), HORIZON EUROPE STRATEGIC PLAN 2021 – 2024, Brussels.

European Commission (2021c), Horizon Europe Programme, Standard Application Form (HE RIA, IA)
Application form (Part A) Project proposal – Technical description (Part B), Version 3.0, Brussels.

European Commission (2018a), A NEW HORIZON FOR EUROPE Impact Assessment of the 9th EU Framework Programme for Research and Innovation, Brussels.

European Commission (2018b), ANNEXES to the Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL establishing Horizon Europe – the Framework Programme for Research and Innovation, laying down its rules for participation and dissemination, COM(2018) 435 final, Annexes 1 to 5, Brussels.

European Commission (2017), Horizon 2020, Work Programme 2016 – 2017, 13. Europe in a changing world – inclusive, innovative and reflective Societies, Brussels.

Jackson, Andrew and Uhrig, Bettina (2021), Impact Strategy Survey Summary and Results, Brussels.
https://www.earma.org/wp-content/uploads/2021/05/Impact-Strategy-Survey.pdf

SwissCore (2020), A new Horizon for Society? Analysing the integration of Responsible Research and Innovation in Horizon Europe, Brussels.
Technical University of Denmark (n.d.), Strategy 2020-2025 Technology for People.

Uhrig, Bettina (2019/20), Impact of Social Sciences and Humanities for a European Research Agenda. Valuation of SSH in Mission-oriented research – Rethinking Societal Impact – Collaboration with Stakeholders, fteval Journal for Research and Technology Policy Evaluation, Issue 48, July 2019, pp 122-126.
https://repository.fteval.at/447/1/Journal48_10.22163_fteval.2019.378.pdf
Republished in: Wissenschaftsmanagement, 02 – 2019, Publikation: Oktober 2020, 24. Jahrgang, ISSN 0947-9546, pp 127-132.

Williams, Kate and Lewis, Jenny M. (2021), Understanding, measuring, and encouraging public policy research impact, Aust J Publ Admin., 80:554-564, https://doi.org/10.1111/1467-8500.12506.

AUTHORS

BETTINA UHRIG
Oslo Metropolitan University (OsloMet), Norwegian Social Research Institute (NOVA)
Postboks 4, St. Olavs plass, 0130 Oslo, Norway
E: bettina.uhrig@oslomet.no
ORCID: 0000-0002-8798-1352

BARBARA SPANÓ
The Ministry of Higher Education and Science, Danish Agency of Higher Education and Science
Haraldsgade 53, 2100 Copenhagen Ø, Denmark
E: bsp@ufm.dk
ORCID: 0000-0001-7295-8638

KEYWORDS: Horizon Europe, Indicators, Impact Pathways, Stakeholders

FTEVAL JOURNAL
FOR RESEARCH AND TECHNOLOGY POLICY EVALUATION

… RTI POLICY EVALUATION AT THE INTERFACE BETWEEN PRACTICE AND ACADEMIA …

The **'fteval Journal for Research and Technology Policy Evaluation'** contributes in a quality assured way to exchange between various stakeholder groups in the area of science, research, technology and innovation policy evaluation. It addresses policy-makers, practitioners, evaluators and the academic community. Thematic editions alternate with thematically open issues. By now 53 issues have been published, which can be accessed and downloaded from the fteval-repository and the fteval homepage. The fteval Journal is open access. The journal and each published paper receive a separate DOI. All paper contributions are quality controlled.

Researchers and practitioners from the RTI policy domain are invited to send contributions to the editor of the **'fteval Journal for Research and Technology Policy Evaluation'** (office@fteval.at). The contributions can be submitted either in German or English.

If you want to receive the **'fteval Journal for Research and Technology Policy Evaluation'** free of charge, please fill out this form and send it to **office@fteval.at** or drop it into the infobox at one of our conferences and events.

first name .....

last name .....

institution .....

address .....

postal code .....

city/place .....

country .....

e-mail .....

By filling in this information you agree to* receive the fteval Journal by mail (free of charge)
* to receive the electronic newsletter of fteval, which informs about pertinent issues of RTI evaluation in Austria and the EU via email (such as calls for contributions for the next fteval Journal) and thus you entitle the fteval to process your indicated personal data for the purpose of transmission of information to you.
You can revoke consent at any time by informing us via eMail (office@fteval.at).

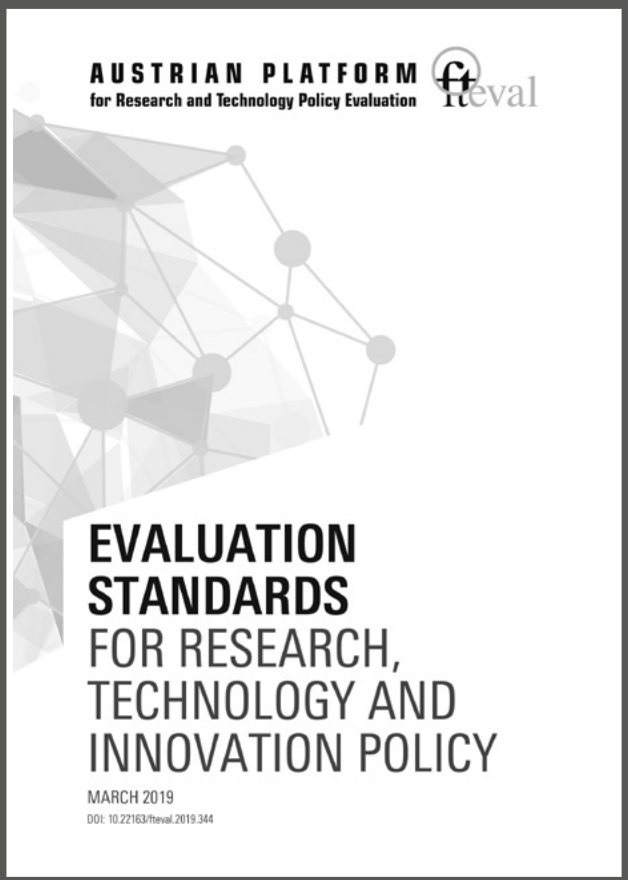
AVAILABLE ONLINE!

LATEST EDITION OF THE

'EVALUATION STANDARDS FOR RESEARCH, TECHNOLOGY AND INNOVATION POLICY'

OF THE AUSTRIAN PLATFORM FOR RESEARCH AND TECHNOLOGY POLICY EVALUATION

Download in German:
fteval.at/content/home/standards/fteval_standards/



Download in English:
fteval.at/content/home/standards/fteval_standards/index.jsp?langId=2

Do you want print copies?
Please send an eMail to office@fteval.at

AUSTRIAN PLATFORM FOR RESEARCH AND TECHNOLOGY POLICY EVALUATION (fteval)

c/o ZSI – Centre for Social Innovation GmbH
Linke Wienzeile 246, 1150 Vienna, Austria
T +43 1 495 04 42 - 33
F +43 1 495 04 42 - 40
E office@fteval.at
W www.fteval.at

DESIGN

W carotte.at

PRINT

Print Alliance

Gedruckt auf PEFC™-zertifiziertem Papier.

AUSTRIAN PLATFORM FOR RESEARCH AND TECHNOLOGY POLICY EVALUATION

members:

Austrian Federal Ministry of Education, Science and Research (BMBWF), Austrian Federal Ministry for Digital and Economic Affairs (BMDW), Austrian Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology (BMK), Austrian Cooperative Research (ACR), Austrian Council for Research and Technology Development, Austrian Institute of Technology (AIT), AQ Austria – Agency for Quality Assurance and Accreditation Austria, Austria Wirtschaftsservice (AWS), Christian Doppler Research Association (CDG), convelop Cooperative Knowledge Design GmbH, Austrian Research Promotion Agency (FFG), Austrian Science Fund (FWF), Institute for Advanced Studies (IHS), Industriewissenschaftliches Institut (IWI), Joanneum Research Forschungsgesellschaft mbH, Austrian Institute for SME Research (KMU Forschung Austria), Ludwig Boltzmann Society (LBG), Agentur für Bildung und Internationalisierung (OeAD), Austrian Academy of Sciences (OEAW), Technopolis Austria, Vienna Business Agency – A service offered by the City of Vienna, Austrian Institute of Economic Research (WIFO), WPZ Research GmbH, Vienna Science and Technology Fund (WWTF), Centre for Social Innovation (ZSI)

AUSTRIAN PLATFORM
for Research and Technology Policy Evaluation

fteval

c/o ZSI – Centre for Social Innovation GmbH
Linke Wienzeile 246, 1150 Vienna, Austria

T +43 1 495 04 42 - 33
F +43 1 495 04 42 - 40
E office@fteval.at
W www.fteval.at

The fteval Journal for Research and Technology Policy Evaluation is a forum to discuss current evaluation practices in the field of research, technology and innovation policy.

ZVR-Zahl: 937261837
ISSN-Nr. 1726-6629
DOI: 10.22163/fteval.2022.537