Fraunhofer ISI

Report on the Impact of Science

Current challenges, concepts, and state of the art in impact assessment

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

Report on the Impact of Science

Current challenges, concepts, and state of the art in impact assessment

Susanne Bührer, Alexander Feidenheimer, Rainer Frietsch,
Maria Karaulova, Ralf Lindner, Torben Schubert

Karlsruhe, March 2023

Compiled on behalf of

Bundesministerium für Bildung und Forschung BMBF

Am Kapelle Ufer 1, 10117 Berlin

Picture credits

Cover page: shutterstock.com/Lucky Step

Recommended citation

Bührer, S.; Feidenheimer, A.; Frietsch, R.; Karaulova, M.; Lindner, R.; Schubert, T. (2023): Report on the Impact of Science - Current challenges, concepts, and state of the art in impact assessment. Berlin/Karlsruhe: BMBF/Fraunhofer ISI.

Published

March 2023

DOI:10.24406/publica-1132

Table of Contents

Figures

1 Zusammenfassung und Handlungsempfehlungen

Im Rahmen des vom BMBF geförderten Vorhabens wurde zunächst eine detaillierte Bestandsaufnahme der akademischen Literatur zum Thema wissenschaftlicher Impact sowie zu existierenden Impactmesssystemen und -methodiken durchgeführt. Aufbauend darauf wurden zwei Workshops mit nationalen und internationalen Expert/-innen durchgeführt, um die Erkenntnisse aus der Literaturrecherche zu spiegeln und kritisch zu diskutieren. Insgesamt hat sich gezeigt, dass ein klarer Bedarf, ja sogar eine Notwendigkeit existiert, dem gesellschaftlichen Impact von Wissenschaft und dem Wissenschaftssystem mehr Aufmerksamkeit zu schenken – mehr als dies in der Vergangenheit in Deutschland der Fall war. Unter wissenschaftlichem Impact verstehen wir den Beitrag, den Forschung für die Gesellschaft leistet. Dies impliziert, dass die Gesellschaft – oder ihre Subsysteme – die Richtungen und Ziele dieser Beiträge definiert. Gesellschaftliche Herausforderungen oder gesellschaftliche Entwicklungsziele könnten einen allgemeinen Rahmen für solche Definitionen bieten.

Mit der Erwartung, dass Wissenschaft Wirkung erzeugt, ergeben sich unmittelbar drei große Herausforderungen. Die erste ist das Attributions-Problem, d. h. die Tatsache, dass Wirkungen in der Regel über verschiedene und auch mehrere Wege entstehen, so dass direkte Ursache-Wirkungsbeziehungen nur schwer zu identifizieren sind und teils nicht einmal sinnvoll definiert werden können. Die zweite Herausforderung besteht in einer adäquaten Messung und Bewertung, für die nicht nur relevante Daten und Indikatoren fehlen, sondern auch ein Verständnis und eine Modellierung von System- Dynamiken. Drittens hängt die zeitliche Verzögerung des Auftretens von Wirkungen mit der ersten Herausforderung der Zurechnung zusammen, führt aber auch zu der Notwendigkeit einer konzeptionellen Trennung von politischen Interventionen und Wirkungserzeugung, die über die üblichen Zeiträume des Politikzyklus hinausgeht.

Es reicht daher nicht aus, dem ohnehin schon zu simplen linearen Modell der Wissenschafts- und Forschungsproduktion, das vom Input über den Output zum Ergebnis und schließlich zur Wirkung führt, am Ende einfach eine weitere Spalte "Impact" hinzuzufügen. Insbesondere die Erwartung, die Wirkung anhand traditioneller oder sogar neu entwickelter Indikatoren zu messen, ist aufgrund der Multikausalität und des verzögerten Eintretens der meisten Effekte zu kurzsichtig. Bei der Festlegung von Wirkungszielen muss beispielsweise berücksichtigt werden, dass verschiedene Arten von Wirkungen – dabei sind wirtschaftliche, wissenschaftlich-technische, ökologische, politische, kulturelle oder gesundheitliche Wirkungen von besonderem Interesse – im Widerspruch zueinander stehen oder zu einem komplexen Trade-off zwischen einer Wirkungsart und einer anderen führen können.

Gesellschaftliche Herausforderungen als Orientierungsrahmen für die Politikgestaltung sind inzwischen weithin akzeptiert. Eine systemische Perspektive wird aus unserer Sicht helfen, diese Herausforderungen zu bewältigen. Damit Wissenschaft und Forschung zur Bewältigung dieser Herausforderungen beitragen und eine gesellschaftliche Wirkung erzielen können, ist ein grundlegender Wandel in der Wissenschaftspolitik und insbesondere in der Wissenschaftsbewertung erforderlich. Die Umsetzung einer solchen Politik ist jedoch nach wie vor eine Herausforderung, die sich in verschiedener Hinsicht manifestiert. Da direkte Kausalitäten und vor allem direkte und zeitnahe Wirkungen nicht erwartet werden können, ist die ex-ante Definition angestrebter Wirkungen sowie die Festlegung von Wirkungszielen unumgänglich, damit das Wissenschaftssystem oder die Wissenschaftsorganisationen ihre Aktivitäten so ausrichten können, dass sie die vereinbarten Ziele erreichen. Ihre Bemühungen, diese Ziele zu erreichen, sind das, was evaluiert und bewertet werden sollte, und nicht die erzeugten Wirkungen selbst. Darüber hinaus sollte die Definition der Wirkungsziele für bestimmte Forschungsorganisationen aufgabenspezifisch sein und sowohl die Governance-Strukturen als auch die disziplinären Unterschiede berücksichtigen.

Um dies zu erreichen, ist die Zusammenarbeit einer Vielzahl von Akteuren und Stakeholdern notwendig, die über heterogene Kompetenzen und Interessen verfügen. Um dieser Heterogenität bei der Politikgestaltung und -bewertung Rechnung zu tragen, sind neue und andere Formen des politischen Engagements erforderlich. Mit einer systemischen Perspektive und ressort- oder institutionenübergreifenden Zuständigkeiten sind gemeinsame Anstrengungen zu unternehmen, bei denen alle Beteiligten – Politik, Verwaltung, Intermediäre sowie Wissenschaftsorganisationen – als verantwortlich für die Bereitstellung von Wegen für gesellschaftliche Wirkungen angesehen werden sollten. Dies wird zu einer erhöhten Relevanz von ressortübergreifenden Themen- und Zieldefinitionen führen. Darüber hinaus werden Fragen der Begleitung von Missionen, anstatt sie nur zu initiieren oder zu finanzieren, an Bedeutung gewinnen, was eine andere strategische Positionierung von Politik und Politikmaßnahmen erfordert. Eine systematische Wirkungsperspektive ist erforderlich, um die Wirkungsprozesse so effektiv wie möglich zu gestalten.

Um also die Wirkung von Wissenschaft auch auf der systemischen Ebene messen und bewerten zu können, müssen neue theoretische und empirische Ansätze entwickelt werden. Wenn Wissenschafts- und Innovationspolitik politische Interventionen gestalten will, die die gesellschaftliche Wirkung erhöhen, muss eine andere Perspektive als die bisherige eingenommen werden. Gleichzeitig gilt es, die konzeptionelle und methodische Basis für das Verständnis und die Messung der Wirkung des Wissenschaftssystems zu verbreitern. Die relevanten Akteure des deutschen Wissenschaftssystems im Allgemeinen und das Bundesministeriums für Bildung und Forschung im Besonderen sollten eine Grundlage dafür schaffen, wie das Wissenschaftssystem und seine Beiträge angesichts globaler Herausforderungen und zunehmender Bedeutung der gesellschaftlichen Wirkungen von

Forschung besser verstanden und bewertet werden kann, um den "neuen Gesellschaftsvertrag für Forschung" zu erfüllen. Der Rückgriff auf alte Konzepte und Messgrößen, so wichtig sie im Einzelnen auch bleiben, reicht nicht mehr aus.

Handlungsempfehlungen:

- Es sollten ein anderer Politikstil und andere Erwartungen in Bezug auf die Erzeugung gesellschaftlicher Impacts festgelegt werden (da diese nicht direkt messbar sind).
- Es ist notwendig, Wirkungen als Prozess und nicht als Ergebnis zu verstehen.
- Folglich sollten die Beteiligten einen Rahmen schaffen, der darauf abzielt, das Engagement für die Erzeugung von Impacts (produktive Interaktionen) zu fördern, anstatt sich nur auf die Wirkungen selbst zu konzentrieren und diese zu messen.
- Wirkungsziele sollten im Vorfeld formuliert und das Wissenschaftssystem bzw. die Wissenschaftsorganisationen hinsichtlich des Engagements zur Erreichung dieser Ziele bewertet werden. Dabei sollte durchaus auch eine Normativität bei der Festlegung der Ziele in Kauf genommen werden.
- Die Definition dieser Ziele sollte sich an bestimmten gesellschaftlichen Herausforderungen und Bedürfnissen orientieren und auf der Grundlage eines gemeinsamen Verständnisses und einer gemeinsamen Zielvereinbarung festgelegt werden. Die Beteiligung von Interessengruppen aus Gesellschaft, Politik und Wissenschaft bei der Festlegung solcher übergeordneten Zielsetzungen sollte angestrebt werden.
- Die Schwierigkeit, Ursache-Wirkungs-Beziehungen bei der Erzeugung von Auswirkungen zu messen, macht die Verteilung von Forschungsmitteln auf der Grundlage früherer Auswirkungen – wie auch immer gemessen – problematisch. Stattdessen sollte die Wirkungsmessung als Prozess verstanden werden, der den beteiligten Akteuren das Lernen erleichtert.

2 Introduction: Current challenges

Public research organisations (PROs), both university and non-university institutions, have the mission to conduct research and to increase a country's knowledge-stock. In Germany, as in most science and innovation oriented countries, the researchers in the science system are committed to scientific rules and regulations, inherent in the science system itself. Above all, they aim to increase the scientific knowledge based on the code of conduct in science. In addition, freedom of science refers to the independence in defining topics and research questions. The corresponding processes are in the spirit of academic self-organization and claim independence from political influence. In sum, freedom of science implies that generating knowledge is the only objective of the science system.

However, all public research organisations are partly or fully funded by national or regional governments and therefore by taxpayer's money. In consequence, both funding organisations and policy makers often demand an efficient and target-oriented research system. Besides, public opinion and discussions in the media often refer to this line of thought, which has become known as "value for money". Mainly representatives from the science system emphasised the antagonism of the "value for money" perspective with the "freedom of science" that would implicitly be harmed by such kinds of expectations. They argue that with this perspective, both the scientific knowledge creation as well as the definition of the research agenda would be subordinate to economically valuable research that is of use to the economic system.

More recently, a discussion with less antagonistic viewpoints on the contributions of PROs to society has taken place. It focuses on mission orientation in (science and innovation) policy making as a new perspective. Next to (only) economic relevance, further goals have been accepted as being at least as relevant as the economic dimension, for example contributions of the science system to addressing societal challenges (environment, health, demographic change, food security, sustainable and affordable energy, transport, communication, etc.), sustainable development goals, or to cultural or societal development in general. While the "value for money" notion directly refers to an input-output relation or efficiency questions, the contributions to missions or societal goals are mainly target questions of effectiveness and – even more importantly – do not refer to short-term outputs, but to effects beyond academia. For these effects, the notion of "societal impact" has become popular, especially in stressing the contrast to simple output-oriented perspectives. Therefore, while the output debate focuses on the scientific system itself, the impact discussion widens the focus to include effects outside the inherent (and expectable) academic effects.

To summarize the above, questions regarding the contribution of the science system to societal goals took centre stage. These contributions, commonly referred to as (societal) impact of science, do not only challenge the well-established habits and frameworks in the science system like the "freedom of science" narrative, but also challenge science and innovation policy making itself. They do this by shifting not only the focal points of funding from input to impact dimensions they also challenge the expectations by funders on the science system. Finally, the evaluation and assessment perspectives as well as the policy processes themselves require adaptations to the newly established goals.

The following paper gives an overview on current challenges, concepts, and state of the art impact assessment. (At least) three overall challenges have to be considered when talking about the impact of science: 1) How to define it? 2) How to measure it? 3) How to reflect on the particular governance framework for science policy in Germany? This report tries to shed light on international discussions and to contribute to a discussion on the (societal) impact of science in Germany.

According to the UK Research Excellence Framework (REF 2012, p. 48), an "impact is defined as an effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia [...]". In evaluation research and practice, a kind of consensus has also emerged according to which an "impact" is a medium- to long-term change; even long-term change that goes beyond the direct beneficiaries of an intervention. This distinguishes "impacts" from "outputs" as directly measurable effects of an intervention that occurs among the direct target groups, and "outcomes" as broader and more medium-term effects among the "beneficiaries". In the case of impacts, it should also be noted that, in addition to the temporal component, a distinction is generally made between direct and indirect as well as intended and unintended effects.

The assessment of impacts is an important function of policy evaluation. It requires measuring change and the attributing the observed changes to the policy. With the rise of the New Public Management, expectations have grown regarding the contribution of science to welfare and society. At the same time, establishing causal relationships between policy interventions and observed changes poses a theoretical challenge as well as empirical and methodological problems. The European Court of Auditors (2008: Annex I) identified i.a. the following challenges:

1. **'attribution problems'**: Due to the large number of external factors (effects of other policies or programs, changes in the legal framework, societal changes, etc.) it is apparently difficult to clearly attribute outputs, not to mention outcomes and impacts, to specific aspects of the RTDI process. It is often not possible to distinguish which effects are directly related to the interventions and its actions. It is also hard to say how these changes, which have occurred as interventions, are implemented in complex environments.

2. **'measurement problems'**:
    - 'data availability': generally speaking, most data (that is available), in particular the data collected through monitoring systems, is on inputs and outputs, rather than on outcomes or impacts.
    - 'understanding the dynamics': given the apparent difficulty to model the relations between inputs and outputs, the inner workings of the research and innovation processes are unclear; and in the absence of an explicit intervention logic, the process of implementation itself is mostly considered a 'black box'.
    - 'comparability of results': the extent to which a comparison of results between different scientific fields can be made is limited. In addition, the evaluation of results is more difficult for basic (or fundamental) research, than for applied research or technological development. The main reason for this is the fact that basic research generally consists of unique, non-replicable procedures. Even in the field of applied research, results vary considerably and cannot therefore be compared easily.
    - 'aggregation': Simply concluding from evaluation results in terms of outputs, outcomes and impacts attained at the lower level for the next higher level is challenging, if not impossible.
    - 'adequacy of indicators': A crucial question is whether evaluators are measuring the right thing. Additionally, a 'measurable' indicator for output does not necessarily permit conclusions on outcomes or impacts to be drawn.
3. **'timing problems'**: Typically, there is a considerable time lag from the research or to the generation of outputs and outcomes, so that impact can be assessed. Data needs to be collected over the long term so that meaningful and robust conclusions can be drawn.

Finally, the German research and innovation system is insofar complex as it is shaped not only by the decision-making processes of the Federal Government, but also by those of the 16 "Länder" (federal states). The government of the respective federal state is responsible for financing research and teaching at public universities that are located in that state. Furthermore, the federal states co-fund the Max Planck Society, the Fraunhofer-Gesellschaft, the Helmholtz Association, and the Leibniz Association. These research-performing organizations (RPOs) cover a broad spectrum from basic to applied research with different disciplinary backgrounds and foci and provide research services like large infrastructures such as libraries. The joint research funding by the Federal Government and the "Länder" is based on the joint task of Article 91b (1), sentence 1 of the Basic Law (Grundgesetz, GG). Under this provision, the Federal Government and the "Länder" may cooperate in cases of superregional importance in the promotion of science, research and teaching. The decision-making body is the Joint Science Conference (GWK, Gemeinsame Wissenschaftskonferenz), which coordinates the research policies between the federal government and the state governments. The Federal Government (Bund) and the Federal States (Laender) spent ca. 3.2 billion euro for the Helmholtz-Association, 1.7 billion for the Max-Planck-Society, 1.2 billion euro for the Leibniz-Association and 0.8 billion euro for the Fraunhofer Society. The Federal Government is responsible for 90% of the basic funds for Helmholtz and Fraunhofer but for 50% for Max-Planck and Leibniz.

In Germany, freedom of science, research and teaching is a fundamental right pursuant to Article 5 GG. Accordingly, universities as well as RPOs have a comparatively high autonomy. The German RPOs administer their processes themselves. They are also rather independent in establishing their strategies. However, this is done in close collaboration with their supervisory boards. Representatives from the federal and state governments are members of these advisory boards. Each RPO has the right to choose the scope and intensity of its policy within the broad framework of the respective state law and the federal German Framework Act.

3 How to define impacts of science

As in almost every discipline beyond natural sciences, a proper definition of the subject under investigation is the crux of the whole work. This applies above all to research on impacts of science. Defining impacts a priori affects not only the conceptualisation and operationalisation of the subject, but also the tools and methods used, and consequently the results and conclusions achieved and drawn. It also frames the whole discussion by determining goals, shifting perspectives, and structuring the arguments used to underpin certain positions. In particular, in terms of societal impacts of scientific work, these kind of dynamics are highly vibrant. This is also reflected in a continuously fragmented discussion, which has been ongoing over the last decade.

At present, various disciplines are contributing to a continuously growing knowledge stock in that regard. These include – among others – economics, sociology, political and social sciences, psychology, and of course numerous cross-disciplinary subfields, e.g. indicator-based science research (science of science), transformative research, innovation research, governance of science, welfare and political economics, behavioural psychology etc. However, only a few contributions are able to provide a more or less holistic picture. At the current stage, it is rather questionable whether this is even possible or desirable as the driving force behind the impact discussion can be reduced to competing political and/or scientific agendas. Hence, finding a single proper definition of impacts of science, especially societal impacts, that satisfies all the different needs, and considers various positions is an enormous endeavour that would expectedly miss the point of actually being an ambivalent subject in a highly fragmented discussion led by competing perspectives. As Feller and Stern (2007) put it: "*No theory exists that can reliably predict which research activities are most likely to lead to scientific advances or to societal benefit.*"

Having said that, the main goal of this section is to provide a broad overview of the most prominent definitions of and perspectives on the societal impact of research. The next chapter will introduce the concepts and methodological approaches derived from these definitions and perspectives.

Before delving into the depths of the impacts of science, some terminological clarifications regarding **social** and **societal** impacts should be provided. The literature lacks consistency in this regard, as an EU expert group also criticizes (European Commission 2018, p. 50): "The very concept of social impact usually used in the literature is problematic and it may refer to a wide range of issues, including policy impact [...]. According to the literature review, specific and commonly used indicators are almost inexistent, or the existing ones are just proposals [...]. Most agencies and models do not consider social impact and policy impact at all. Sometimes, they mention them as criteria to be considered, but without specific indicators"

(European Commission 2018, p. 50). Various authors from academia (Bornmann 2013, Holmberg et al. 2019) highlight that the terms are often used synonymously and interchangeably. Holmberg emphasizes that "social impacts" "often refer to a more personal level of influence, affecting people directly or indirectly (Vanclay et al. 2015)". In the EU, on the other hand, a distinction has been established according to which "social impacts" have a relatively comprehensive claim and include economic impact, societal impact, cultural impact, environmental impact, human rights and partly also policy impacts (European Commission 2005; 2017). Societal impacts, meanwhile aim more at the aspects of society in the narrower sense, i.e. argue rather the other way round.

In the following, we will use the term 'societal impacts' as a comprehensive paraphrase describing any kind of effects resulting from scientific work.

It should not be overlooked that societal impacts – regardless of the actual definition used – are always closely connected to their scientific field. The impacts of social sciences differ from the impacts of natural sciences. Societal outcomes of astrophysics are very likely to be substantially different to societal outcomes of medicine and research on infectious diseases. Thus, the more specific a definition of impacts is formulated, the more it is biased towards certain scientific areas. We aim to understand societal impact on a higher and more generic level. However, we acknowledge the fact that this is not always feasible and that definitions have to be adapted to the specific field or area of interest. It is not possible to present all possible definitions as this goes far beyond the scope of this report. We will therefore try to focus on the more generic definition type.

One of the earliest definitions of impacts in current discussions has been provided by Spaapen et al. (2011) within an EU funded project called SIAMPI. This work focuses on assessing societal impacts through the study of productive interactions. The authors argue that this is the most important mechanism through which research leads to socially relevant application. For them, "*social impact of scientific research refers to measurable effects of the work of a research group or program or a research funding instrument in a relevant social domain. The effect regards the human well being ('quality of life') and or the social relations between people or organizations*" (Spaapen et al., 2011). While this definition ought to be as generic as possible, it still introduces several restrictions. On the one hand, impacts are described as a measurable effect with regard to human wellbeing and social relations. On the other hand, the analytical level is narrowed down to a group, program, or a specific funding instrument that produces impacts and a relevant social domain where the effects can be observed. Based on this definition, one might expect a fundamental shortcoming, which is a lack of comparability between different research groups, programs etc. However, the authors have found a rather good approach to avoid this shortcoming with a generic measurement that satisfies the need for comparability sufficiently by focusing on impact generating mechanisms (Joly et al., 2015; Joly and Matt, 2017).

Another perspective on societal impact has been proposed by Bozeman and Sarewitz (2011). The authors argue that the central aspect of science policy-making are public values and funding programmes are primarily focused on achieving them. Therefore, funding programmes should be assessed with respect to ‘public value success/failure’, which is very closely related to the economic terminology of market success/failure. The authors find that although no adequate assessment tools are available, policy-makers will still make choices about funding priorities. Therefore, their focus lies on developing a decision-making tool that is introduced as the ‘public value mapping’ (PVM) model. Bozeman (2007) defines public value as following: *‘‘A society’s ‘‘public values’’ are those providing normative consensus about (1) the rights, benefits, and prerogatives to which citizens should (and should not) be entitled; (2) the obligations of citizens to society, the state and one another; (3) and the principles on which governments and policies should be based’’*. Furthermore, Bozeman and Sarewitz (2011) add that "*public values failure occurs when neither the market nor public sector provides goods and services required to achieve public values*". Although the authors do not focus on science impact specifically, we can still deduce a definition of a generic type from a political decision making perspective. That would be the contribution of science to public values success. An important aspect of this perspective is that generating impacts is primarily seen as an obligation of science policy-making rather than science itself. An almost unique point of view that clearly keeps scientific standards of autonomy and freedom untouched.

A more process-oriented understanding of societal impact is based on the broad categorization of stages leading towards impacts as well as the actual impacts themselves (Buxton, 2011; Buxton and Hanney, 1996; Donovan and Hanney, 2011). Such perspectives do not offer a descriptive definition of the subject under investigation, but rather its conceptualisation. Thus, we will address them in more detail in the next chapter. However, it should be highlighted that processes and the timeliness of impacts are a crucial aspect in most of such concepts. Currently, they are often referred to as impact pathways.

A policy-oriented perspective on impacts can be found in descriptions and definitions provided by ministries, funding bodies, or research councils. We will refrain from discussing all of them, but just mention the most prominent ones. For example, the Economic and Social Research Council (ESRC) in the UK defines research impact as "*the demonstrable contribution that excellent research makes to society and the economy*".[1] Furthermore, they differentiate between two dimensions of impact, namely academic impact (we might also call this dimension scientific impact) as well as economic and social impact. Other approaches – e.g.

1 http://www.esrc.ac.uk/research/impact-toolkit/what-is-impact/

the Commonwealth Scientific and Industrial Research Organisation CSIRO (2015) – differentiate between societal, economic and environmental impact and exclude the academic area explicitly. The same goes for the Research Excellence Framework (REF) – a periodic research evaluation of higher education institutions in Great Britain – which defines impact as "*an effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia*".[2] Finally, in the principles of impact evaluation, the OECD points out that "*[...] impact evaluation is an assessment of how the intervention being evaluated affects outcomes, whether these effects are intended or unintended.*"[3] By defining impacts as effects that are either intended, or unintended, the OECD reveals one of the most challenging parts of societal impact, especially with regard to a standardised evaluation of it. Moreover, it also highlights the difficulties in the demand for accountability of research funding instruments.

The closing remarks on this chapter shall touch on the current situation in Germany. In Germany – compared to other nations like the Netherlands, Great Britain, Australia, or Sweden – the impact agenda is less visible in policy documents. Nonetheless, there are some programmes and interventions that are meant to facilitate processes which can be considered as relevant for generating (societal) impact of research and innovation (R&I). The most prominent programme to support research and innovation in Germany is the "Hightech-Strategy" which follows a mission-oriented approach similar to the societal grand challenges of the EU[4]. By establishing thematic priorities, the federal government and the Federal Ministry of Research and Education (BMBF) in particular, aim to improve the quality of life in Germany in specific areas through advancements in knowledge production and innovation. The main priorities in the current round of the programme (Hightech Strategy 2025; HTS 2025) are located in the following six areas:

- Health and healthcare;
- Sustainability, energy and climate;
- Mobility;
- Urban and rural development;
- Security;
- Economy and working environment 4.0.

However, the predefined missions of the HTS 2025 are only one of three pillars of the new research and innovation policy in Germany. The second one shall establish a solid basis for future developments by strengthening national competencies in technological progress,

---

2 https://re.ukri.org/research/ref-impact/

3 http://www.oecd.org/dac/evaluation/dcdndep/37671602.pdf

4 https://www.bmbf.de/en/the-new-high-tech-strategy-2322.html

human capital and skills, and by fostering a participative society. The third one aims to ensure a culture that is open to innovation and risk-taking while focusing on generating impacts out of knowledge production, strengthening the entrepreneurial spirit, and expanding knowledge and innovation networks.

Hence, the HTS 2025 increasingly brings the impact of R&I on the agenda of R&I policy compared to previous HTS programmes. The aspects concerning the generation of impacts out of knowledge production are particularly intriguing. However, its detailed implementation process reveals that this part is primarily focused on knowledge and technology transfer, i.e. the attempt to bring more ideas/inventions to the market and thus produce more innovations. Thus, Germany follows a very market-oriented impact strategy assuming that the market is the focal point of impact generation and most societal benefits are resulting from market related activities.

3.1 Our definition and approach

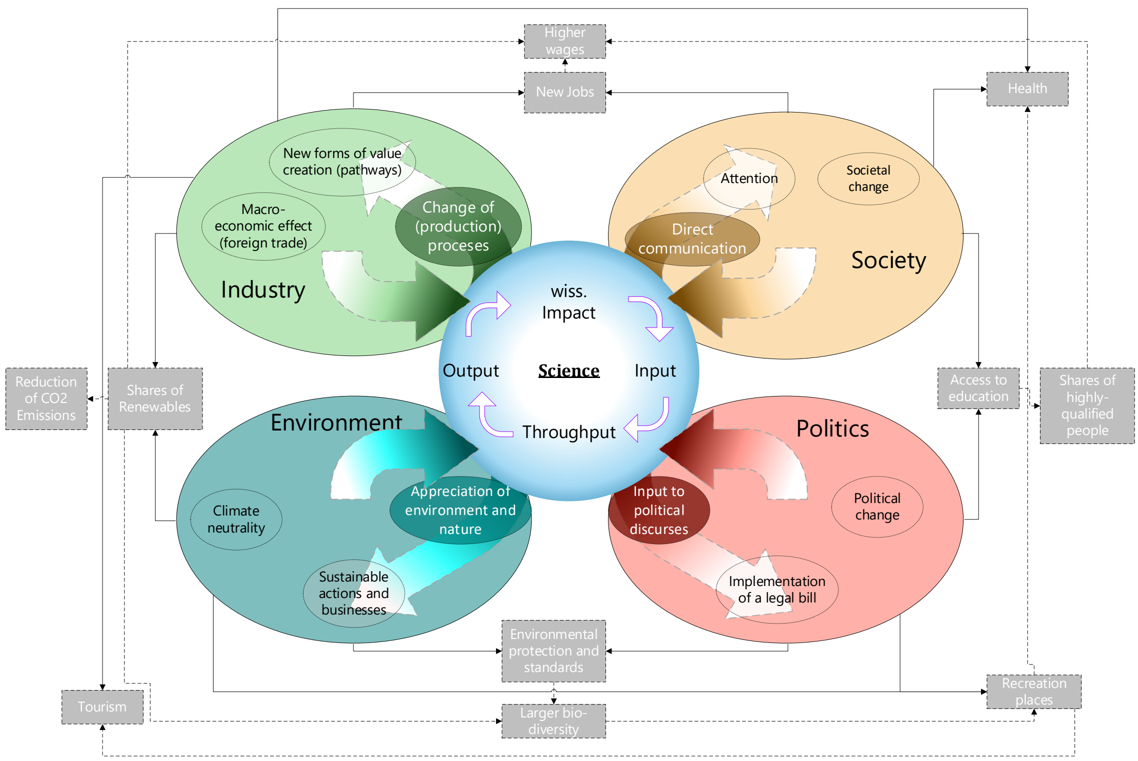
A simple and intuitive definition is to think of societal impact as "the demonstrable contribution that [...] research makes to society and the economy"[5].

It is important to emphasize – and also to consider empirically and conceptually – that impact is not simply a relabelling of output or outcome. Impact means a result that has an outward effect. Impact occurs when either system boundaries are exceeded and effects take place outside the scientific system or when outcomes significantly influence the scientific system itself.

Accordingly, we propose a definition of impacts that relies on social systems theory to understand society as a whole as consisting of (interacting) subsystems (e.g. science, environment, economy, politics, social system). Figure 1 shows the two types of impacts (crossing the boundaries of the science system and fundamentally changing the science system itself) from this system perspective.

5 http://www.esrc.ac.uk/research/impact-toolkit/what-is-impact/

Figure 1: Schematic representation of societal impact



Source: Own representation

Since we speak of societal impact, our definition explicitly addresses the impact on different parts of society. This can include scientific results that individually or collectively have economic impacts (e.g. jobs, competitiveness,...), policy impacts (new policies, laws, regulations,...), environmental impacts (reduced CO2 emissions, higher share of renewable energy, ...), health impacts (fewer hospitalizations, longer life expectancy, ...), impacts on society (less inequality, better knowledge and skill levels, better understanding of science in society, ...), impacts on culture (changing values, habits, attitudes, ...) and also impacts on science (e.g. outputs/results leading to new knowledge elsewhere, which spread across disciplines or extend beyond institutional borders, ...).

However, "...impact is a complex and multi-faceted phenomenon and should be understood as a non-linear, network-oriented process" (LERU 2018).

Since impact sometimes occurs with a massive delay, clear causal effects or strictly targeted political or strategic interventions can hardly be assessed. Nevertheless, there are various conceptual and methodical approaches to addressing this problem in the sense of tracing a well-founded and empirically comprehensible contribution. In science, for example, the concept of "productive interactions"[6] has attracted great interest in order to conceptualize impact. This approach focuses on the process of creating impacts and aims to increase the possibilities and probabilities for impacts to occur. We see this concept as a worthwhile approach for the German science organizations and the science system in Germany as a whole.

6 See for example: Spaapen, J. and van Drooge, L. (2011). Introducing 'productive interactions' in social impact assessment. Research Evaluation, 20(3), 211–218; Joly, P.-B. and Matt, M. (2017). Towards a new generation of research impact assessment approaches. The Journal of Technology Transfer, 1(4), 537; Joly, P.-B., Gaunand, A., Colinet, L., Larédo, P., Lemarié, S., and Matt, M. (2015). ASIRPA: A comprehensive theory-based approach to assessing the societal impacts of a research organization. Research Evaluation, 24(4), 440–453; Molas-Gallart, J.; Tang, P. (2011): Tracing 'productive interactions' to identify social impacts: an example from the social sciences. In: Research Evaluation, 20 (3), p. 219–226.

4 Conceptual approaches

4.1 National evaluation programs

One of the best-known research evaluation systems is the UK's former Research Assessment Exercise (RAE), which was established by the University Grants Committee (UGC) in the mid 1980's. The main purpose of this program was a meaningful distribution of public funding among universities. The underlying approach of this program was a call for submissions of scientific output, e.g. publications, and subsequently a review of the work received. Therefore, the assessment was based on a conventional ex post evaluation (Hill, 2016). However, this funding allocation system has been criticised by many researchers as rather discriminatory because it has ranked disciplines instead of institutions (Hare, 2003). Furthermore, the focus of most contributions was to submit high excellence journal articles. Other important aspects like knowledge dissemination and transfer, generating ideas as well as teaching were left unrecognised (Bence and Oppenheim, 2005). Despite all the negative outcomes, the UK continued this programme while they tried to learn from each evaluation round that they performed. The last RAE was held in 2008, after which it was replaced by the Research Exercise Framework (REF), that was adopted in 2014 for the first time.

The replacement of RAE by REF was not only driven by the bad image and criticism experienced over the years but also by a substantial shift in knowledge production from *mode1* – the traditional theoretical, or experimental science – to *mode 2*, which is comparatively more application-oriented, trans-disciplinary and socially distributed (Nowotny et al., 2003, 2008). This change illustrates the demands facing science in addressing societal challenges or problems during their research (Hill, 2016; Matt et al., 2017; Raftery et al., 2016).

By implementing this very idea into the evaluation of higher education institutes, the societal impact agenda was brought to life. In respect thereof, it was suggested that the Higher Education Funding Council for England (HEFCE) funds research that is closer to the needs of end-users and stakeholders from industry. Following this recommendation, the HEFCE began to ask scientists to illustrate the 'pathways' their research might take towards impact (Martin, 2011). In the first round of REF, the assessment of research impact became an inherent part of it, thereby accounting for 20% of the overall score (REF, 2012, 2014).

The UK's interest in evaluating the research impact was not unique. In fact, the initial idea came from Australia's Research Quality Framework (RQF), which was already appointed in 2004 by the Minister of Science, Education, and Training in order to forward the innovation strategy focusing on the broader economic, social, and environmental benefits for industry and business (Donovan, 2008). Today, societal impact in Australia is evaluated by The

Commonwealth Scientific and Industrial Research Organisation (CSIRO). CSIRO’s strategy in evaluating the impact of research is based on a mixed-methods approach. The impact types (economic, environmental, and social) are assessed by using monetised, quantitative, and qualitative data. Thus, the evaluation approaches have a wide range like the cost-benefit analysis, non-market valuation, non-monetary quantification, as well as qualitative methods. In general, all evaluations serve the purpose of addressing the five A’s: accountability, allocation, analysis and advocacy (CSIRO, 2015). However, the underlying framework described by Morgan (2014) as a ‘pathway to impact’ follows a logic model with the following stages: inputs, activities, outputs, outcomes, and impact, which is very similar to the payback framework introduced by Buxton and Hanney (1996) for assessing the ‘payback’ of health-related research.

Among the EU member states, one of the most developed evaluations of societal impact is implemented within the Dutch Standard Evaluation Protocol (SEP) (Bornmann, 2013; Donovan, 2008) . The evaluation is performed periodically on each research unit's responsibility, or group within each research institution, and is assessed afterwards by an independent committee. Besides the self-assessment, the committee is provided with additional documents regarding the research unit and holds interviews with heads of the units. With respect to the evaluation, SEP is focusing on three key assessment criteria of research. First, the research quality captures the contribution to the scientific knowledge by looking at e.g. scientific publications. Second, the research’s relevance to society is assessed by quality, scope and relevance of contributions made to economic, social or cultural target groups. Third, the research’s viability captures the strategy to meet its goals – scientific as well as societal – in the future (KNAW et al., 2014). Here, the assessment of the societal impact of research is captured within the second point that refers to societal relevance. The evaluation of the societal relevance is further described within a guide provided by a collaborative project of multiple Dutch organisations called Evaluating Research in Context (ERiC, 2010). The evaluation approach of ERiC is very closely related to the SIAMPI model of Spaapen and van Drooge (2011) that focuses on investigating productive interactions of researchers with relevant research stakeholders.

Other EU member states also have societal impact assessments implemented into the evaluation of publicly funded research programmes. However, in most cases, these are less explicit and visible compared to the approaches described above. In Germany for example, it is anchored within section 7 of the Bundeshaushaltsordnung (BHO; Federal Budget Regulation) which enacts the principles of efficiency, economy as well as cost and performance accounting for public expenditure. This also includes a respective performance control in order to ensure the underlying principles. Furthermore, the performance control is based on evaluations that should also consider planned and unplanned effects of the performed activity (Carlsson et al., 2002; Engels, 2013; Federal Republic of Germany, 2017).

Although section 7 of the BHO does not specify the effects in more detail, the resulting evaluations mostly consider a mission oriented assessment approach (Böttcher et al., 2014). The underlying missions are predefined by policy makers within Germany's High-Tech Strategy that has the ambitious goal of solving societal problems through technological development (BMBF, 2014, 2017). A study funded by the EC, The Evaluation Partnership (TEP) and Centre for European Policy Studies (TEP and CEPS, 2010) found that none of the 27 EU member states have a comprehensive impact assessment tool yet, but rather integrated evaluation mechanisms focusing on specific impact areas (e.g. economy or health). However, 21 of the EU member states follow a respective integrated approach.

4.2 (Semi-) qualitative evaluation approaches

Productive interactions (SIAMPI)

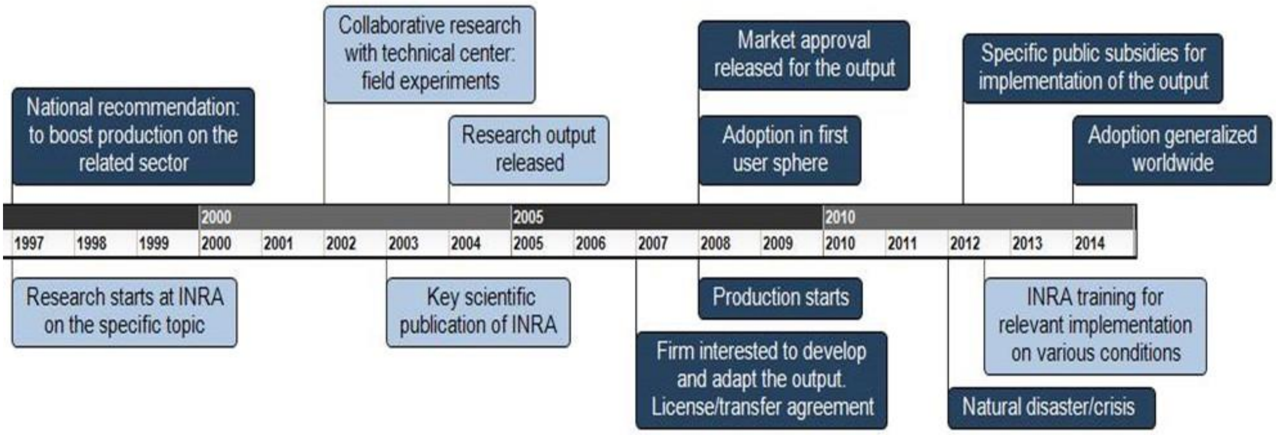
The SIAMPI project (Social Impact Assessment Methods for research and funding instruments through the study of Productive Interactions between science and society), funded by the EU Commission, has focused on the process of knowledge production and the interactions that take place there between the participants as the central "social impact" (cf. Spaapen and van Drooge 2011; Siampi final report). The participants include the researchers themselves as well as initially unspecified stakeholders. According to the study authors, an interaction is considered productive when the knowledge generated is applied or used. In essence, the result of a productive interaction is then understood as a change in behaviour (Spaapen and van Drooge 2011, p. 212). A total of three possible forms of interaction with stakeholders are differentiated, i.e., the environment of the researchers: (1) Direct interactions, e.g., in person, by telephone, etc., operationalized, among others, by the number of researchers holding multiple positions, membership in advisory boards, or presentations to lay audiences; (2) indirect interactions, e.g., through texts or other forms, operationalized by a quantitative indicator called CRA (contextual response analysis), which can be used to measure interest in certain reports and papers (on the Internet, via keyword searches); and (3) financial interactions, e.g., through research contracts etc., operationalized as contracts, licenses, project funds, shared infrastructure, industry-funded PhDs, etc. (Spaapen & van Drooge 2011, pp. 213, 217).

Thus, this approach does not examine factors such as quality of life, social security, etc. (which can be considered social impacts), but rather the impact of research on society, insofar as it can be observed as a change in behaviour (Siampi Final Report, p. 1f.). It is emphasized that social impacts cannot always be clearly distinguished from other impact dimensions. The authors see the most likely points of contact with economic and cultural factors, less so with technical and environmental effects (Siampi Final Report, p. 5f.).

Ideal-type impact pathways (ASIRPA)

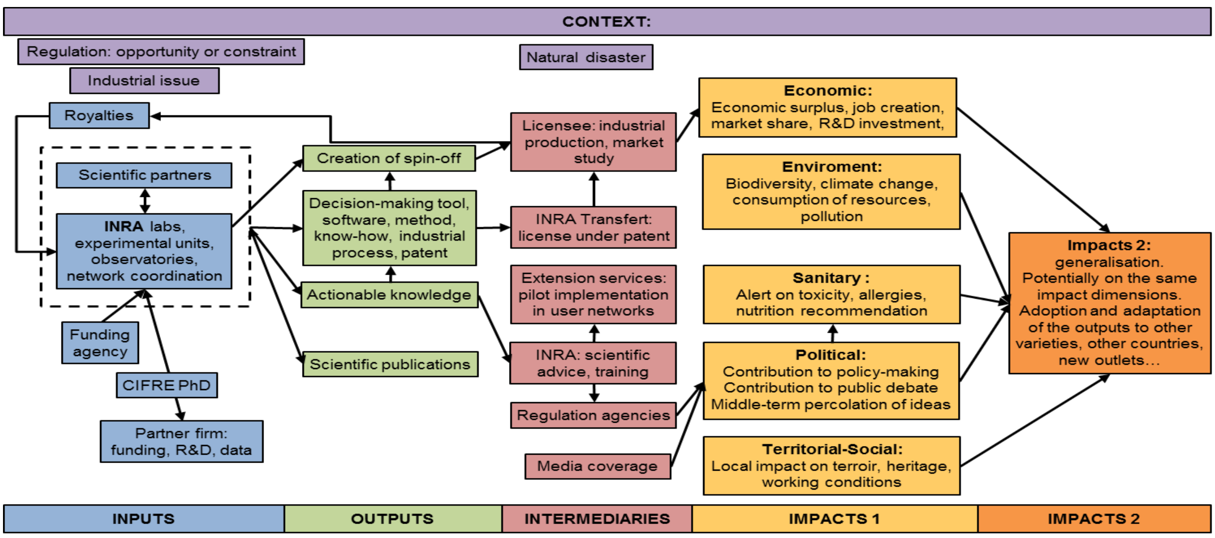
ASIRPA investigated the impact of the research portfolio of a French research organization in the field of agricultural research (INRA) and produced a large number of standardized case studies (Joly et al., 2015). The occurrence of research results in final products and processes was investigated, backtracking to research activities as well as an analysis of the involved actor networks and impact mechanisms was performed, and finally aggregation and typing were completed. The following figures illustrate the core elements of this approach:

Figure 2: Hypothetical chronology by ASIRPA



Source: Joly et al. (2015, p. 446)

Figure 3: Impact pathway example



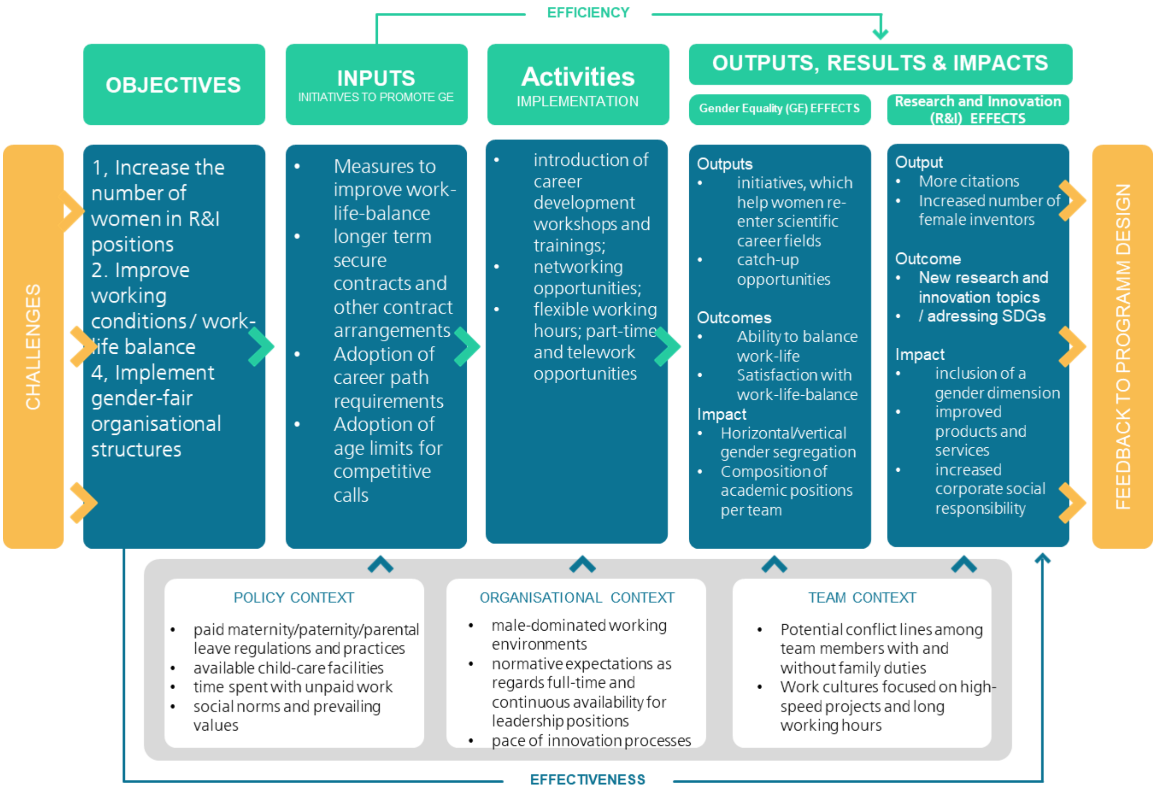
Source: Joly et al. (2015, p. 447)

Theory based impact evaluation (TBIE)

One of the most promising approaches to address these challenges in general and in particular to solve the "black-box" problem of evaluations (Astbury and Leeuw 2010) are the approaches of theory-based evaluation (TBE) or theory-based impact evaluation (TBIE) (see Leeuw 2003, 2021, Nonie subgroup 2 Impact Evaluation). This is a heuristic that examines change primarily in terms of which factors have contributed to an observed change, without claiming to delineate or model explicit cause-and-effect relationships in a separable way. The focus of TBIEs is on "why and how"? Focusing the question on "why does something work?" implies elaborating a theoretical understanding of how something should logically work to achieve the desired outcome, i.e. an intervention theory. On the other hand, as Bloch et al. (2014, p. 106) put it, "[...] given a focus on informing future use, the question of why and how the impacts were achieved are as equally important as the question of what the impact itself was."

The theory-based impact evaluation approach (Kalpazidou Schmidt and Graversen 2020) uses "theories of change (ToC)" as a model for the way in which change is expected to occur, and ideally it also provides evidence for how the effects came about (Mayne and Johnson 2015). They are similar to the "logic model" often used in development interventions that depicts an undertaking, program, or project in a brief, visual format (McLaughlin and Jordan, 2004; Knowlton and Phillips, 2012), but explicitly include a reflection on the assumptions underlying an evaluation. Mayne and Johnson (2015) state that a theory of change develops a framework that indicates how an intervention will work. Thus, a tested and verified ToC can be the starting point for evaluating the intervention's or program's contribution to measured effects, and is consistent with a growing strand of research that discusses how a theory of change can contribute to the evaluation and understanding of policy interventions (Funnel and Rogers 2014; Rogers 2008). A graphic representation of how to implement the above requirements for modern impact measurement is provided by the theory-based evaluation concept EFFORTI, which has its origins in looking at gender equality interventions but whose basic structure can be applied in a variety of contexts (https://efforti.eu/efforti-toolbox-intro).

Figure 4: Example of a theory-based and context-sensitive impact model

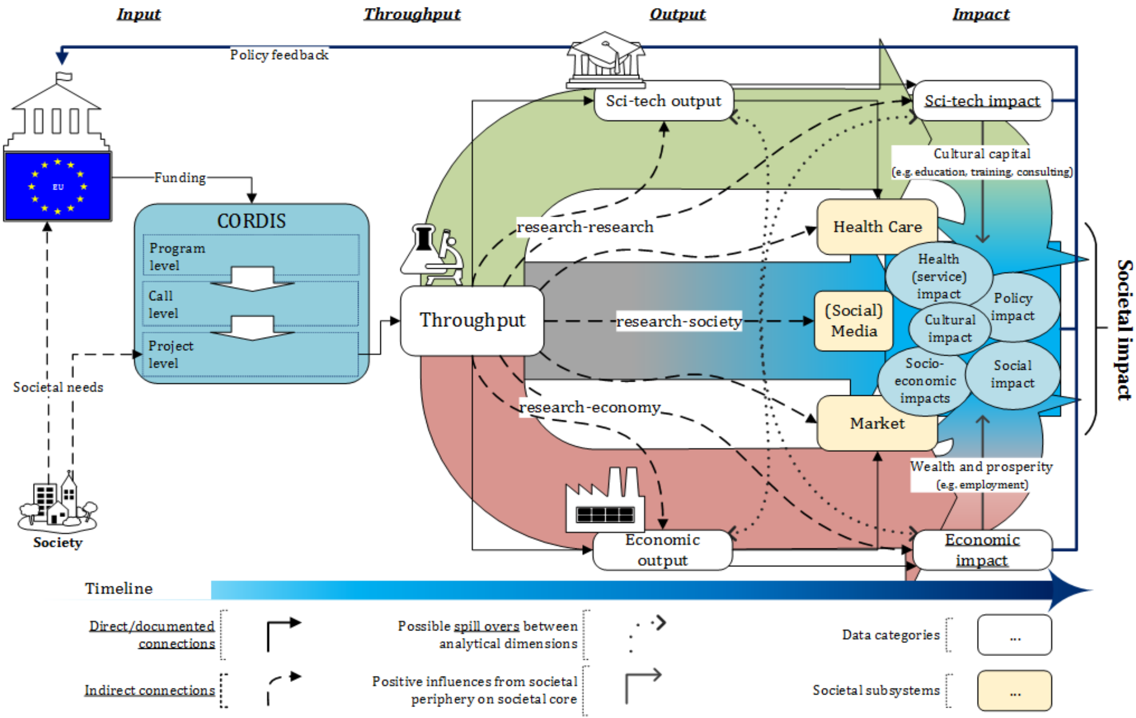


Source: Own representation.

4.3 Linear concepts

A more data-driven assessment concept that is based on a linear logic model has been developed by Feidenheimer et al. (2019); it is shown in Figure 5. Although it is applied to evaluating the societal impact of health related research funded by the EC, the model is also adaptable to other national or private funders, funding as such, or – considering small modifications – the R&I system on the national/ international level as a whole. The model is divided into four analytical stages: the well-known input, throughput and output stage and finally the rather new impact stage. Each stage is represented by characteristic data categories/clusters (white rounded rectangle) with a description of the information type they contain. The data that has been collected and analysed refer to funded projects where R&I is performed. Therefore, the lowest level of analysis is the project level. The different clusters are connected by at least one common information entailed. In all, four possible connections represented by distinctive arrows can be made by analysing the data in each cluster. The black solid arrows indicate a direct/documented connection, e.g. explicit references, citations, occurrences, etc. The dashed arrows represent an indirect connection, e.g. information/data addressing common topics without directly referring to each other. Further, the dotted arrows symbolise the interconnectivity between data clusters that are separated by different analytical perspectives/ dimensions, e.g. potential spill overs from academic output to economic impact. The grey arrows indicate translational activities or beneficial influence of separated impact categories on the society additionally enhancing the societal impact. In the background, three different ways of generating impact are illustrated by green (scientific technological impact path), red (economic impact path), and grey-blue (direct dissemination activities and topics being addressed by research and discussed within the society) coloured arrows. Finally, the actual societal impact is placed within a box between both peripheral dimensions, i.e. the scientific-technological impact with science as the main driving force and the economic impact with technology, respectively. In the centre, various possible impact fields (e.g. health impact, policy impact, etc.) are mentioned exemplary.

Figure 5: A model of societal impact assessment (AMOSIA); H2020 - Data4Impact



Source: Feidenheimer et al. 2019

Furthermore, the timeliness of the whole process is illustrated by the timeline below the graphical representation of the model. Therefore, the analytical stages as well as respective data categories are arranged according to the time course. However, one has to be aware that this is not an exact science. On the one hand, we can assume that dissemination and utilization of knowledge (measured by output indicators) can only be performed after the knowledge has been generated (measured by throughput indicators). On the other hand, it is not appropriate to assume that economic impact occurs simultaneously with the academic impact. Thus, the timeline is only applicable to the horizontal axis, but not to the vertical. Furthermore, the timeline does not represent a strict sequencing of all processes, as the stages are rather a heuristic help to structure the analysis. This means that a particular process does not need to be finalised before a subsequent restart. Instead, they all proceed partly simultaneously with time.

The scientific-technological impact describes a demonstrable contribution of R&I made to academia or in other words, the consequences of knowledge dissemination for the academic system and scientific practice. The economic impact, on the other hand, describes the consequences of knowledge utilisation within the economic environment. In this case, one can think of exploitation of new markets, enhanced production efficiency and productivity, lower production cost and ultimately, growing competitiveness. Simultaneously, both dimensions affect the society resulting in a societal impact. The societal benefits occur rather indirectly and gradually in a systematic and geographical sense than directly and overarchingly. The societal consequences of economic activities are mediated through the market where the demand and supply side come together. Society benefits from new or cheaper products that might simplify everyday life and enhance the general wellbeing. The economy receives monetary benefits in turn. This interaction can be facilitated and strengthened by the media through advertising, or weakened by a negative report about the company (e.g. concerning irresponsible resource extraction and production). Furthermore, the academic dimension is directly linked to the societal dimension through institutions, e.g. hospitals, universities, museums, and other research infrastructures that have a direct link to the public sphere. Here, professionals are putting their new knowledge into practice directly affecting society. Finally, societal impact can also be achieved by communicating the results of a particular research activity directly (e.g. through social media).

5 Measurement: Data sources and indicators

A wide range of approaches to measure the impact of R&I have been developed and used over the last thirty years, each of them relying on different assumptions and measuring impact in different ways. We will now broadly review and classify the most frequently used approaches into those relying on indicators and those relying on economic/econometric modelling seeking to measure the economic impact of R&I.

5.1 Indicator-based approaches

Innovation is defined as the process from the idea to the market (OECD and Eurostat, 2005). Modern innovation theory distinguishes between three paradigms with different processes and different characteristics (Sundbo, 1991, 1995, 1998). The first is the classical Schumpeterian paradigm where the entrepreneur is the central actor in the innovation process, uniting the financial, intellectual/innovative, and economic responsibility in one person. The second paradigm focuses on the R&D expenditure and new technological achievements, mainly in large multinational companies. This second paradigm was the dominant mode of the innovation process after the Second World War. It is still partly in play in several sectors, but was recently substituted by the third paradigm, where R&D still plays an important role, but knowledge creation as well as successful market activities are not only based on R&D expenditures. New factors gain more importance, among them marketing, networking and especially the interaction between technologies and services. More recently new dimensions and new criteria for innovation are discussed, for example the implications of the digitalisation of the industrial production. Besides, and this is also a result of the new perspective on innovation in the third paradigm, participation and transparency as well as societal impact receive much more attention nowadays. Innovations are not only assessed by their economic effects, but even more so by their contribution to overall societal goals.

Measuring and assessing the impact of science and innovation policies is even harder to address. While the potential outreach of innovation policy becomes more and more limited the closer this process approaches the market – this was already pointed out in the early 2000s in a seminal work by Kuhlmann and Arnold (2001) – recent analysis in empirical innovation research have tried to address exactly this issue.

Because of these developments, there is an increasing demand for additional evidence for the performance of the research and innovation systems, and the societal impact/benefit they bring. This creates a demand for new indicators and data/evidence. However, there are caveats involved.

It is not surprising that the European Commission's key monitoring frameworks (e.g. the Innovation Union Scoreboard; H2020 key performance indicators) have been focusing on the inputs, outputs and results of the programmes, while programme impacts have been mostly assessed qualitatively.

While the traditional and well-established innovation indicators mainly address input and throughput – both on the micro and macro-economic level (Crépon et al. 1998; Frietsch et al. 2015; Frietsch and Schmoch, 2006; Neuhäusler et al. 2016; Ophem et al. 2002; Rammer et al. 2015) – have proved their predictive power and validity, especially for technical product innovations (this was the main focus of the second paradigm), output and impact indicators are scarce. Several approaches have been proposed to extend the list of available indicators and to overcome the existing gap. Processes beyond formal R&D as well as service innovations have been in the scope since the late 1990s, but next to micro-economic data based on innovation surveys, the list of established indicators is rather limited, even after almost 20 years. Some conceptual and also empirical progress has been made in the service research (Gallouj and Gallouj 2000; Gallouj and Windrum 2009; Gallouj 1997; Gallouj 2002; Hauknes 1998; Hipp and Bouncken 2009; Tidd and Hull 2003) and in knowledge-intensive businesses (Biege et al. 2013; Coombs 2003; Gotsch and Hipp 2012; Muller and Doloreux 2009; Shearmur and Doloreux 2013). Macroeconomic indicators similar to R&D expenditures, patents or publications are not yet established. Trademarks have been used as a throughput indicator similar to patents in the context of services (Gatrell and Ceh 2003; Gotsch and Hipp 2012; Mendonca et al. 2004; Sandner and Block 2011; Schmoch 2003), but neither their data quality and predictive power, nor their theoretical/conceptual integration have reached the level that patents play for technological product and process innovations. On the output side, knowledge intensive businesses are seen as complementary to high-tech sectors. On the sectoral level it is the turnover, value added or production that are used as output indicators. However, while high-tech sectors are defined by their R&D intensity (share of R&D expenditures over turnover or production value), knowledge intensive businesses are defined by the formal qualifications of the employees. Again, the theoretical/conceptual relation of knowledge and innovation is more indirect than between R&D and (technological product and process) innovation, even though knowledge might be the most important source of innovation, both in services and in manufacturing.

When it comes to measuring the impact of innovation activities and innovation policy, the theoretical/conceptual framework as well as the available indicators are largely missing. Some first steps have been made based on qualitative approaches (Daimer et al. 2012; Kuhlmann, Rip 2014; Lindner et al. 2016), mainly in evaluation studies, where qualitative programme aims were linked to qualitative (and sometimes also quantitative) input and throughput data. Other attempts have taken place in the British Research Exercise Framework REF (previously RAE), and have also been proposed by the Swedish Research Council for the national evaluation model (SRC, 2015). Using qualitative approaches could be helpful to evaluate deeper aspects of specific cases, but it does not offer easy steps to compare distinct areas in a systematic way.

5.1.1 Patent indicators

Patents are among the most important indicators for the output of R&D processes and are frequently used to assess the technological performance of firms, technology fields and economies as a whole (Freeman, 1982; Grupp, 1998). A large amount of patents thus indicates strong efforts in R&D activities and therefore a higher innovative output. However, large patent portfolios are also strategically useful, for example, to block competitors in the same or adjacent technological areas or prevent especially smaller potential competitors from entering relevant markets (Blind et al. 2006; Blind et al. 2009; Neuhäusler, 2012).

Besides the indicator function of patent filings as such, patents bear a large amount of additional information, e.g. references to previous patents and scientific publications or the outcome of the examination process. This information can be used by researchers to indicate for example the development of new technological trajectories, the pace of technology cycles, the linkage of a patent to science or the economic and technological quality of patents.

Patent data analysis is booming nowadays, increasing the body of literature in the field – and as the literature grows, so do new insights and new knowledge as well. Not all analyses that use patents apply the same methods and definitions (Moed et al. 2004). First and foremost, patents can be seen and analysed from different angles and with different aims: the technological view allows prior art searches or the description of the status of a technology; micro-economic perspectives – for example – allow for the evaluation of individual patents or the role of patent portfolios in technology-based companies; a macro-economic angle offers an assessment of the technological output of national innovation systems, especially in high-tech areas.

Starting from a simple legal perspective, patents give an exclusive right of usage to the applicant for a limited period. In addition, patents can be interpreted as an indicator of the codified knowledge of enterprises, and, in a wider perspective, of countries. As an innovation indicator, patents fit into a system of further indicators to describe scientific and technological competitiveness and to analyse innovation systems. This means, in the input-throughput-output indicator model presented above, patents are used as an output of R&D processes and are therefore classified as throughput. They are not themselves a means for economic success, but can be an important input to future economic activities that rely on the patented technology. R&D processes can either be measured by the input – for example, expenditures or human capital – or by the output. In order to achieve a more precise approximation of the "black box" (Schmoch and Hinze 2004) of R&D activities both perspectives, i.e. input and output, are needed.

From the perspective of innovation systems, patents indicate the output of technology generating processes and thereby enable the assessment of the technological competitiveness of nations. Especially international patent filings are meaningful for comparisons, as they reflect activities in international markets where national and multinational companies meet with their competitors directly and on neutral ground.

5.1.2 Bibliometrics – publication and citation based indicators

The science system is an important part of any national innovation system. It does not only provides basic research results that feed into further innovation processes in public research institutions and especially in enterprises. It also provides scientific knowledge that is of the utmost importance for innovation activities in general. Scientific knowledge generated in institutional as well as project-funded research activities, mainly public research organisations – be it universities or non-university institutions – is crucial input to the further innovation process. The creation, diffusion and application of scientific knowledge are essential foundations of the technological activities and also key elements in the performance of a national innovation system. Basic scientific research plays a significant role in technological development. First of all, industrial capabilities and competencies clearly rely on highly skilled personnel as part of the major source of innovation, which heavily depends on the success of training and education during the scientific research. Also, scientific achievements are definitely substantial bases of technological development and innovation.

Furthermore, scientific knowledge is generated in scientific discourses. This means being in exchange with fellow scientists and receiving peer criticism. These scientific discourses take place for example by scientific journal publications, which can be read and cited by others and used for one's own research. The "currency" of this reception and acknowledgement are citations/references that are made to previously published work. Bibliometric data analyses make use of these citations and do not only take into account the quantity or the bibliographic information – such as institution, author, country or field of research – but also these citations and interpret them as a sign of visibility or even of quality of the cited publications.

The use of bibliometrics to measure and assess the performance of scientists has meanwhile become a well-established standard. It also has become an important instrument for innovation systems analysis and for the assessment of the scientific and technological competitiveness of nations. Publications are used as indicators of the scientific strength of science systems or as of early-stage innovation activities and may offer a perspective on future innovations, based on current scientific and technological strengths.

Bibliometric studies make use of scientific publications and their citations to measure the performance and impact of national science systems, parts of national innovation systems. One of the underlying assumptions is that (scientific) impact can be measured by the number of citations since scientific achievements are mostly published in journals, so that other scientists have access to them and can consequently them.

In sum, bibliometrics are the scientific measurement of scientific output by means of publication data. Bibliometrics mostly deal with publications in scientific journals. However, recently also books and conference proceedings have gained more and more interest, although this area is still too young and methodological restrictions prevent a wider use. In this latter context, no standard analysis has been established yet, particularly as the databases covering these kinds of publications are still under development.

5.1.3 Alternative indicators for the R&I impact assessment

In addition to bibliometric or patent indicators, a number of alternative indicators measuring research output are in use. Following the notion that the new public management paradigm in science has increasingly emphasized competitive elements of governance, competitively acquired third party funds have found wide recognition. The rationale is that third party funds reflect research activities, where the underlying research proposals were subject to selection processes and eventually succeeded over competing proposals. According to that logic, a research organization acquiring many third party funds presumably performs better in terms of research. In several countries including Germany, indicators relating to third party funds are treated as a measure of research output and have a dominating role in the performance-oriented distribution of state funds for R&I (Schmoch and Schubert 2009). However, several studies have questioned the usability of third party funds. They have provided evidence that the relationship between the share of third party funds in total financing and genuine measures of research output – including bibliometric indicators – is non-linear (Jansen et al. 2007, Schmoch et al. 2010). Thus, there appears to be a threat of over-reliance on third party funds, which is detrimental rather than beneficial for research output. A further argument against the use of third party funds is that they represent an input rather than an output, as the mere acquisition of a project does not yet imply any output.

A further indicator, which has been used for measuring research output, is based on counting doctoral graduates (compare Jansen et al. 2007). This indicator as well as third party funds are intensively used in practical evaluations of research output in many official indicator systems maintained by European governments. It also shares a key conceptual problem with third party funds, as doctoral graduates can be considered both input into and output from research activities. Furthermore, it is unclear whether the necessary teaching activities to produce doctoral graduates do not render the indicator for teaching rather than for research.

The potential of **altmetrics** to identify the impact of research is highlighted when surveying the societal impacts of research. Experiences made within the British REF (Research Excellence Framework) show that scientific "output" is often correlated with the "impact" measured with the help of altmetrics, i.e. that conventional scientific excellence and the impact of research can be mutually dependent (Wilsdon et al. 2015, p. 48). The idea to use altmetrics in addition to classical bibliometric methods is grounded not only in the observation that "even where government documents, for example, quote academic work these references are not citations in the traditional sense and are therefore not picked up by bibliometric analysis. **Grey literature produced by academics tends to be more used by policymakers but its impact is difficult to capture**. Firstly citations are not made in the usual way and secondly academics have been slow to realise the importance of using tagging information such as DOIs in order to allow these references to be tracked" (Wilsdon et al. 2015, p. 45).

5.2 Economic approaches

The class of economic approaches for the impact assessment of R&I consists of two basic models. The first one tries to monetarize the value of R&I. Such approaches allow for, at least in principle, measuring of the returns on investment. The second approach tries to determine the impact on economic figures, such as the growth of companies, which collaborate with R&I organisations. Studies trying to assess the economic value of R&D in monetary or economic terms follow different methods, but arguably, the most widespread ones are the "input-output approach" and the "production-function approach".

5.2.1 The input-output approach

A wide range of studies – mostly for individual universities – have focused on monetarizing the values of the R&I organizations by using statistical input-output-tables. The general idea behind this approach is to regard the economic activities of the focal organizations as a reflection of their economic value. For example, the capital investments made by a university are perceived as an increase in demand for capital goods and therefore lead to additional economic activities on the side of the suppliers. Likewise, a university attracts students and employs staff, which leads to additional consumption expenditures in the region. These types of economic expenditures are available from official university statistics. In addition, these types of studies typically also account for multiplier effects resulting from the demand increases for consumption and investment goods, by allowing the consumer producers and the supplier in turn need to raise their demand for goods in order to serve the additional demand induced by the R&I organizations. These multipliers can be calculated either through a Keynesian multiplier methodology or based on regional statistical input-output-tables. Such analyses therefore allow estimating the economic return including "second round" effects. They also allow for a distinction between direct effects of the expenditures spent by universities and employment effects as well as indirect multiplier effects (for an overview compare Glorius and Schultz 2002; Schubert et al. 2012). Typical estimates of the multipliers range between 1.2 and 2.2 implying that one Euro spent on by the university leads to an additional spending in the economy of 0.2 to 1.2 Euro (compare Glorius and Schultz 2002; Blume and Fromm 1999). Due to data limitations and the complexity of the approach most studies have focused on just one university. A notable study measuring the economic impact of more than a single organization was provided by Glückler et al. (2015), who found that the nine full universities in the German state of Baden-Württemberg induced an increase in value added of about € 3.7 bn. This corresponds to a multiplier of 1.8 with respect to the funding provided by the state. While extensively used, the biggest drawback of input-output or Keynesian models is that they focus only on effects induced via the increase of demand for investment or consumption goods. These studies therefore do not take into account the knowledge effects of R&I (Glückler et al. 2015; Schubert and Kroll 2016). Thus implicitly the role of R&I organizations is displaced to that of a regular tangible infrastructure investment, such as construction. Arguably, however, the real purpose of R&I organizations is to generate and diffuse new knowledge. Measuring the knowledge effects (sometimes referred to as supply-side effects) is practically impossible using input-output or Keynesian demand side models. Methods, which aim at measuring the knowledge-related supply side effects, typically build on the notion of the production function.

5.2.2 The production function approach

The production-function approach (PFA) to innovation is conceptually rooted in macroeconomic growth theory. As the founding father of neoclassical growth-theory, Solow (1957) showed that the accumulation of physical capital could only account for short-run but not for long-run growth. This is due to decreasing returns to scale to capital accumulation. Instead, only if the total factor productivity (TFP) increases, there would be positive long-term growth. Although Solow made a case that the origins of TFP-growth lie in technological progress, the neoclassical growth model does not explain technological progress as an outcome of economic activities. Rather, it takes the rate of technological change as exogenously given. Thus, the model leaves the technological change essentially unexplained.

It seems intuitive that the central economic activities undergirding technological progress are related to research and development (R&D), which firms engage in to maximize their dynamic profit rate. While later works culminate in the development of endogenous growth theory (Romer 1990, Aghion and Howitt 1990) show that purposeful and deliberate R&D activities can indeed sustain long-term growth. Earlier works were primarily empirical in nature and sought to determine the economic value of R&D. An important approach to estimate the economic value of R&D follows the PFA, which consists of three important methodological elements

1. *the output*, which is typically scalar and can, depending on the level of analysis, reflect some level of economic output (e.g. value added), performance (e.g. firm-level productivity) or knowledge (e.g. patents),
2. *the inputs*, which are typically multidimensional and include regular inputs (e.g. capital and labour) on the one hand and innovation-related inputs (e.g. R&D) on the other,
3. *a techno-economic law* according to which inputs are transformed into the output.

While substantial literature exists on the effects of universities, only few analyses exist analysing the effects of extra-university research organizations including research infrastructures. A few authors have applied the concept to the case of extra-university public research organizations. Robin and Schubert (2013) as well as Kaiser and Kuhn (2012) analyse the effect of public research organizations on firm performance in terms of product innovation, patenting and productivity of public research organizations in general. Nonetheless, neither of these authors are able to differentiate the effects into university and extra-university public research organizations. Comin et al. (2019) are the first to present analyses using the PFA for the specific case of the Fraunhofer-Gesellschaft. They show that for the firms interacting with Fraunhofer through research contracts firm performance is substantially driven by the interactions. In particular, 22% of the turnover of growth of interacting firms can be explained by the interaction. For productivity growth, the respective figure is 11%.

Though under restrictive assumptions the results on the firm level can be used to determine the macro-economic returns (Robin and Schubert 2013). However, little is known about impact of public research organizations on the economy as a whole. Recently a few studies have emerged which try to identify the macroeconomic returns based on the PFA. Instead of starting from the firm level and then aggregating up the results at the macroeconomic level, these studies rely on regional macroeconomic data. Conceptually, a generic version of PFA still applies, but instead of using it on the level of the firm, the regression model is specified on the level of the region. While coarser in nature, these studies evade the problem that a science-industry collaboration may positively affect the focal firm but negatively affect the focal firm's competitors. In addition, the PFA on the regional level is better apt to identify knowledge spill overs. The earliest application is Goldstein and Renault (2004) who investigate the role of universities in the USA by estimating inasmuch they affect the regional development in the 312 main statistical areas (MSA). They find significantly positive effects. Extending the regional PFA, Schubert and Kroll (2016) analyse the effects of universities on the GDP in Germany.

A further variant of the production-function approach has relied on time-series modelling with non-stationary data. This approach makes use of complex econometric methodologies to identify the effects of spending on public R&I on productivity. These analyses usually rely on country-level official statistical data, which is available for long time series.

A final approach is the so-called micro-to-macro approach as used for example by Allan et al. (2021) who analyse the economic effects of the Fraunhofer-Gesellschaft on GDP, investment, employment and sectoral composition based on a Computable General Equilibrium (CGE) model used on macroeconomics. Such models characterize an economy by a set macro-economic general equilibrium conditions and then model how certain shocks (here an increase in Fraunhofer budgets) would affect the overall economic equilibrium. While CGE-models are principally pure macro-type of models, the model used by Allan et al. (2021) is micro-routed because it relies on model parameters that are derived from external econometric results coming from quantitative modelling approaches of behaviours at sub-macro levels. A related example of such approaches is found in Comin and Giarda (2019), who use a Dynamic Stochastic General Equilibrium model, where key parameters where estimated based on linking Fraunhofer contract data to firm level data from the German part of the Community Innovation Survey (Comin et al. 2019).

5.3 EU-impact pathways in H2020 and Horizon Europe (HE)

As mentioned above, more recent impact assessment approaches take account of the fact that an observed intervention is often only one factor among many that contribute to a change, and that it is therefore of fundamental importance to consider the obstructive and supportive context factors in the impact models or Theories of Change. Given this background, these more recent approaches often prefer to talk about a **contribution** rather than an attribution.

Current examples of applying impact pathways include the European Commission's approaches to assessing the impacts of the European Framework Programmes (Bruno and Kadunc 2018; 2019) and the impact pathways to survey the effects of research infrastructures (**RI Paths**).

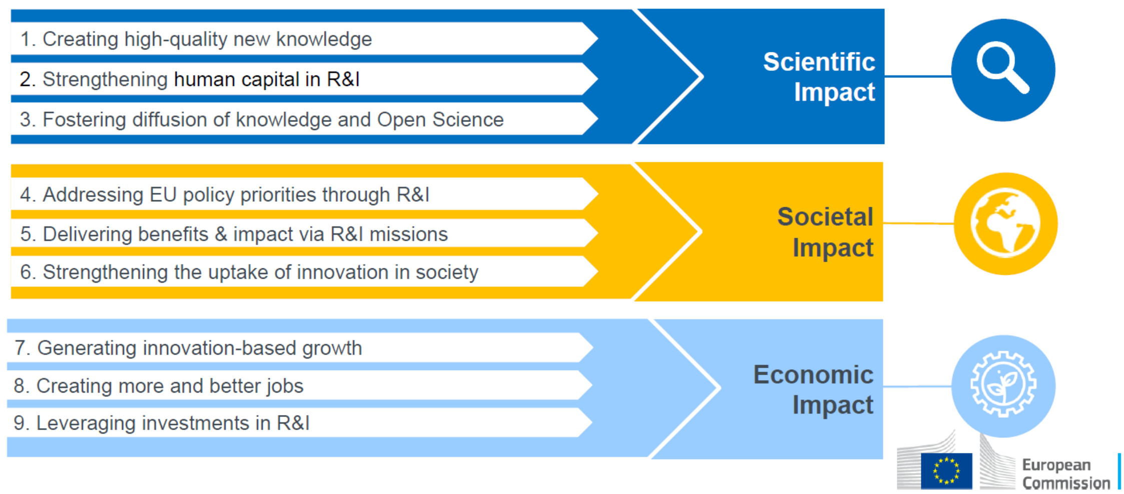
In recent years, the European Commission has launched a number of projects and studies with the intention of measuring impacts. Several of these projects offer both a conceptual framework for impact evaluations in general and empirical approaches for measuring them. Although many of these projects address impacts in specific areas (e.g. health) or at an organizational or individual level, they address a broad array of impact dimensions, e.g. economic, social, technological/scientific, environmental impact, etc.

The European Commission's **key impact pathways** described in the programme's documents for the planned Horizon Europe Framework Programme are primarily aimed at the following: (1) To tell a "story" about the framework programmes as a whole, according to their respectively defined goals (similar to the "narratives" that are becoming increasingly relevant in applied innovation research). (2) To develop realistic indicators, in order to map short-, medium- and long-term progress. (3) To reduce the efforts for those being evaluated after project completion. (4) To allow for a distinction between management/process and performance indicators.

The nature and characteristics of impacts and impact assessments have been mentioned above. In addition to the topicality, it is above all the challenge of the causality assessment, which provides additional complexity both conceptually and empirically. The Key Impact Pathways approach addresses these types of questions based on five key principles, which they call the PATHS principles (Proximity, Attribution, Traceability, Holism and Stability).

Against this background, the authors from the European Commission have suggested nine key impact pathways including concrete indicators, three for each area of scientific impacts: economic, technological, and societal impacts. **Societal impacts** are defined here as follows: "Societal impacts: Strengthen the impact of research and innovation in developing, supporting and implementing **EU policies**, and support the **uptake of innovative solutions** in industry and society to address **global challenges**" (Bruno & Kadunc, 2018, p. 9).

Figure 6: EU's impact pathways



Source: Bruno & Kadunc (2018): Paving the pathway to impact in Horizon Europe

The Commission had tendered a project that aims at providing a performance-based monitoring system based on the so-defined multidimensional impact assessment dimensions – the so-called Key Impact Pathways – of the new Framework Programme Horizon Europe.

These nine indicators, grouped in three Key Impact Pathways, built the framework and the structure of the annual monitoring of future Framework Programmes. It can be seen as a performance-based monitoring approach that allows the assessment of progress of Horizon Europe towards its objectives.

5.4 Qualitative approaches: Case studies and formative assessments

Molas-Gallart (2015; 2012) points out that when assessing the impacts (or public value) of research, it is important to consider what the goals of the assessment are. He differentiates three goals of evaluation: allocating funding, control and learning, and improving practices (Molas-Gallart 2015, p. 117). If the assessment is to allocate funding, he believes indicators that are easy to collect and compare are required, while formative goals are more concerned with considering details and contextual factors, as is possible, for example, when using a **case study approach** (Molas-Gallart 2015, p. 118). In concrete terms, he elaborates on the necessary methodological approach of a **context-sensitive, formative assessment** of societal impacts: "A focus on processes requires the use of research methods capable of providing a fine-grained **understanding of the ways in which research generates public value**, and the contextual factors that facilitate or hinder such processes. This work is very time consuming (and therefore expensive), and yields a type of result (awareness of the importance of specific, context-dependent factors) that is useful to improve practice, to adapt policies, and to tweak implementation routines, but can only provide contributory help to make decisions on the distribution of resources" (Molas-Gallart 2015, p. 121f.).

Wilsdon et al. 2015 also highlight the suitability of case study approaches, as these can be used to illustrate **complex interrelationships** and it is not a coincidence that these were also applied within the framework of ASIRPA: "Others also argue that academics who create impact by building long-lasting partnerships with groups and organisations will not be able to demonstrate the depth and detail of impacts through metrics or data alone. A key concern from some critics is that impact metrics focus on what is measurable at the expense of what is important (Wilsdon et al. 2015, p. 44)". At the same time, however, they point out that there are voices in academia that view case studies as "fairy tales of influence" and, against this background, tend to advocate the use of indicators (Wilsdon et al. 2015, p. 49). In addition to the positively rated case study approach, the superiority of **peer reviews** compared to indicator systems is also emphasized: "No set of numbers, however broad, is likely to be able to capture the multifaceted and nuanced judgements on the quality of research outputs that the REF process currently provides. (Wilsdon et al. 2015, p. 136).

6 Insights from two workshops

In view of the difficulties to grasp (societal) impacts in the sense of a long-term change and to attribute this to a certain impulse – a research result, a promotion (attribution problem)), those approaches that focus on the conditions of emergence of societal impacts (especially productive interactions) are of particular importance. There is also a consensus in the relevant literature: "[...] there is an issue where substantial consensus is emerging: the importance of social interactions as the main foundation through which research results find applications that generate social value" (Molas-Gallart 2015, p. 120).

The following section shows that these reflections are in line with current debates on impact research and practice as discussed at the two impact conferences organised with support of the BMBF in May 2021 and February 2022.

In the following, we will present a summary of the discussions during these impact-conferences, which took place in an online-format, with 50-60 participants, each. Among the participants were scholars from the field of science and innovation studies, representatives from research funding agencies including ministries and internationally leading experts in impact assessment. This outlays ongoing discussions about potentials, pitfalls and challenges in impact measurement and evaluation. It does so by first delineating common themes shared by the participants of the workshop. Where possible synthetic lines of agreement were identified. Apparent opposing or conflicting viewpoints are also highlighted in this summary. The overall aim of this summary is therefore to contribute to a summative understanding of the facets of the current debate not only on the impact of science showing commonalities but also on the lines of controversy and thereby generating an understanding for the future of this topic both from a practical and from a conceptual perspective.

From the first workshop, three major themes emerged. Firstly, there were intensive discussions about the reasons for measuring impact at all. Among the participants, the commonalities were relatively pronounced by admitting that impact assessment can strengthen both accountability and systemic learning. Secondly, viewpoints on appropriate methods including their advantages and disadvantages on impact measurement were exchanged. While all participants agreed that there is a need for mixed-method approaches, existing methodologies differ between retrospective and prospective approaches. Among the participants, there were proponents of both approaches. Nonetheless, all participants agreed that prospective and retrospective methodologies share distinct advantages and disadvantages implying that depending on the application both may prove useful.

Thirdly, the effects of the nature and purpose of the evaluation on scientific performance were addressed. Here, a clear agreement emerged that future impact evaluations should put more emphasis on learning. That does not imply that accountability considerations will or should become irrelevant. They are, however more appropriately addressed within the frameworks of regular output evaluations.

6.1 Reasons for measuring scientific impact

Measuring scientific impact can be desirable for a number of reasons. Traditionally and following the logic of new public management, impact measurement has been justified with reference to the concept of accountability. Holding science organizations accountable for their outcomes is seen as a necessary condition for granting them more flexibility in terms of their input use. Output accountability, however, requires at the very least that output is measured and benchmarked against societal expectations. What differentiates impact assessment from simple measurement of scientific outputs (such as publications and the generation of human capital) is that more distant societal effects ("impacts") need to be accounted for. Beyond creating accountability, it should be noted that the output transparency associated with accountability can also be leveraged for lobbying purposes: If research-performing organizations or policy-makers are able to demonstrate the relevance of scientific research, this knowledge can create powerful arguments for a steady stream of funds to science. It can also more generally contribute to creating positive attitudes towards science. In that respect, the workshop participants agreed that impact measurement can also lay the basis for effective lobbying for science. Two more aspects speaking in favour of a consistent impact measurement relate to organizational and systemic learning and the creation of insight into how science, society, policy, and the economy interact to create societal impact. Thus, impact measurement can also open up social learning potentials. While this goal is certainly among the most desirable of the outcomes, it is difficult to achieve. Using impact measurement for organizational and policy learning requires not only a solid measurement framework to create a useful informational basis, but also requires a systemic and reflective approach to strategizing in science organizations as well as policy-making. The necessary strategic capability would in many respects have to be built up. Because of these high demands, it is therefore not surprising that the justification for impact measurement in practice still focuses on allowing for improved accountability rather than on the generation of learning potentials.

6.2 Measurement approaches

Measuring impact rather than direct outputs of science brings about a number of conceptual and empirical challenges. In most cases, the "distance" between cause (e.g. a specific scientific artefact) and the societal effect (e.g. economic, social or political) is usually large. This implies that it is often difficult to establish the causality between them empirically. It is often aggravated by the existence of considerable time lags, which renders the choice of an adequate period difficult. Moreover, the large distance between cause and effect raises questions about correct attribution resulting from the existence of many potentially overlaying parallel processes. Finally, because impact does not only consider economic aspects, which are often easier to measure, it also refers to intangible, cultural and social domains for which reliable data does not often exist.

Existing measurement approaches accordingly differ substantially depending on the specific objective. Quantitative studies relying on econometric and macroeconomic modelling approaches have dominated the estimation of economic impacts (Carree et al. 2014, Schubert et al. 2013, Comin et al. 2018, Roy et al. 2021, Schubert 2021, Allan et al. 2022, Bilsen et al. 2018). Quantitative approaches based on indicators have been extensively used to measure the impacts on technological development (Frietsch et al. 2017). Some few quantitative studies even tried estimating societal impacts such as the emergence of clinical guidelines resulting from research in medicine (Feidenheimer et al. 2018).

Nonetheless, the ambiguities surrounding the definition, delimitation and identification of impacts in a broader sense have rendered quantitative studies to be infeasible. Qualitative approaches have therefore been substantially more important in impact assessments than they have been in evaluating scientific outputs. The role of narratives, the definition of impact pathways as well as case studies therefore come to the fore. Indeed, many of the assessment frameworks outlined by the Research Excellence Framework (REF) invest substantially into understanding the pathways between cause and effects on a theoretical level giving guidance into their measurement. The ASIRPA-model (Joly et al. 2015), a pioneering example for impact research, is a case-based measurement approach, which solves the time-lag problem by selecting concluded cases and tracking them back over time. The attribution problem is approached by defining impact pathways and the lack of data on qualitative impact dimensions is approached by resorting to expert assessments on qualitative scales. While ASIRPA is largely retrospective, a more prospective approach goes into a different direction by defining necessary conditions for impact generation. The theoretical assumption that impact occurs through "productive interactions" between heterogeneous actors seems to be particularly interesting.

Thus instead of waiting for impacts that occur over a long time in the future, which would make timely studies impossible, impact assessment could instead focus on identifying such interactions and take their existence or absence as an indication of whether impacts are likely to occur later on (Spaapen & van Drooge 2011).

Among the participants of the conference, there emerged an agreement that to model and measure impacts more comprehensively, mixed-method-approaches are necessary, because only they are able to combine the empirical and conceptual strengths by employing hard quantitative techniques whenever possible but resorting to qualitative approaches when measurement and attribution problems become rampant. Examples can be found in the impact assessment approach (the Key Impact Pathways) of Horizon Europe, which makes a distinction between short, medium and long-term impacts. In fact, as a rule of thumb it may be stated that the more diffused the impact dimension is, the further it is in the future and the more complex the transmission channels linking cause and effect are, the more desirable it is to resort to qualitative rather than quantitative techniques.

6.3 Dare more normative values!

The second workshop was mainly dedicated to the perspectives from the German science system and the implications of societal impact demands to the actors in the science system and the innovation system as a whole. Accordingly, participants were mainly representatives from different science organizations as well as intermediaries and additional stakeholders such as the BMBF, the German Science Council (Wissenschaftsrat), or the German Science Foundation (DFG).

Starting from the results of the discussions at the first workshop, the focus was directed to two main dimensions, namely the question of how to specifically measure the achievements of German research performing organizations as well as how to achieve/ formulate the expectations of what impact means and how it should be addressed (value-driven system).

An impulse presentation and the subsequent discussion revealed five types or categories of measurements and assessments of the science system, derived from historical developments. The first category that occurred in the 1970s did not consider the effects at all, but was mainly investment-driven, which resulted in a strict input-orientation. The second type of science evaluation shifted the attention to the outputs, mainly driven by fiscal scarcities. Here, efficiency criteria and "value for money" – this is the extreme simplification of the question what taxpayers get (and might expect to receive) for the public investments into science and research – were introduced. The third type is characterized by a further differentiation of outputs and a move towards outcomes as directed/expected outputs instead of general outputs. For example, instead of simply maximizing the numbers of patent applications or scientific publications, directions of the outcomes e.g. in certain areas of particular economic or societal interest were formulated and also their input into further activities like commercialization or licensing were taken into account. The fourth category is characterized by cause and effect relationships, which allow a first glimpse at societal impacts of science, which are not yet exactly impacts, given that the selection of the impact dimensions/contributions are not yet socially negotiated (not yet societal challenges). A fifth category addresses the measurement of the performance of the science system, most often by KPIs (Key Performance Indicators), which happen to lead to "window dressing", negative control effects, and often also to an overestimation of the controllability of science systems. In some cases, the compliance with the KPIs is valued higher than the outputs or outcomes themselves, which might also become a hampering factor for creativity – the core asset of science, research and innovation. Due to the attribution problem, the illusion of effect control by indicators might appear.

Based on the measurement and assessment challenges, a discussion on the specific formulation and selection of impacts for science emerged. The core question in this respect was, if and how a value-driven funding and science system can be established. However, the debate mainly centered around the source and setting of these values, i.e. the normativity of the selection process as opposed to scientific market forces. The need for a value-based instead of a pure efficiency or performance-based system stems on the one hand, from potential lock-in effects, and system failures, on the other hand. In addition, failures to properly measure (reliability) the performance as well as negative effects of the measurement itself hamper these systems. It was stressed, however, that the current science system is under massive pressure to justify itself with respect to its performance.

The science system is driven by a commitment to trueness and veracity, which is both a chance and a duty. This commitment – it was even called 'verocraty', which could be translated as the government or control of veracity – might fail and needs normativity, which could be described as a value-based directionality in order to avoid these failures. Responsible research and innovation (RRI) is such a normative framework that is already established and might be linked to the impact setting as well. In addition, the normativity might be put into the science system from outside, for example by societal development goals and the formulation of expectations to address these.

However, societal development goals or the different challenges that societies face might not be clear and explicit and they might also not be without contradiction. A more challenging issue was raised further in the course of the discussion, if the science system might be held responsible for things that it cannot control or even influence.

Instead of performance measures that might satisfy justification demands (value for money), trust in the science system should guide the funding decisions, especially against the background that different societal changes have different time horizons. The majority of the participants agreed that more normative values should be dared.

It was the opinion of several participants that the aims of a strategic debate on the societal impact of science should pay more attention to the processes of science and of knowledge generation instead of output-oriented frameworks.

When it comes to timelines of politics and policies in relation to impacts, a discrepancy becomes obvious. However, if a shift towards impact assessments of the actors in the science system or the system as a whole is to be implemented, definitions of successful completions might become necessary from policy makers' perspective. Several representatives of the science system stressed that a process instead of an output or performance perspective might solve this problem and satisfy the needs of the political system as well. A debate arose, as to whether a focus on processes is possible without measurement and control. The answer was yes, on the one hand due to joint definitions of the processes. On the other hand, also, the process and its partial steps can be monitored and assessed, but a focus on the outputs or outcomes is not necessary for this. If this is not sufficient, it was also suggested to define impact scenarios instead of individual process-oriented measures and assessments.

6.4 Implications for research evaluation

The workshop participants agreed that evaluation practices need to adapt as the science and innovation systems change. Indeed, as the demand to assess societal impacts of science more broadly increases, evaluation practices must adapt to new realities as well. Specifically, while conceptual and measurement issue prevent a short-termed identification of cause and effect, there is a need to understand impact as a process rather than as an outcome. This shift from an outcome to a process perspective implies not only that measurement approaches are changing towards more qualitative and, if quantitative, to more indirect measures of preconditions. It also means that the purpose of impact evaluations needs to change.

Specifically, if impacts are difficult to observe or if there remains a lack of clarity about their causes, aiming for accountability becomes less relevant. In particular, if research-performing organizations have limited power to influence later impact generation, it is hard to hold them accountable whether impacts occur or not. Rather, observing that impacts did or did not occur gives extremely valuable information about the functioning of the impact pathways. Thus, instead of focusing on measuring and comparing impact performance between organizations, there should be a focus on understanding and improving impact pathways along the whole chain of impacts. Such an endeavour is much more related to processes of collaborative and reflective learning than it is related to invention-creating, performance-based budget allocation, as the accountability aspect would suggest. Impact measurement therefore requires a radical shift in the evaluation philosophy towards learning-oriented approaches. In that respect, impact assessment is not selling the old wine of output measurement in new bottles.

The main implications of the discussions at the second workshop are that an impact perspective is directly connected to the question of setting and monitoring these impacts, which results in a demand of normativity. Societal impacts are derived from societal needs and societal challenges, but it needs to be pinned down to usable and reasonable expectations. The definition, setting and monitoring of these expectations in terms of impact achievements is a joint endeavour of all participants in the system. This, however, should be as free as possible from particular interests, which can be achieved by multiple actors and the building of trust and trustful relationships among all actors in the science system – policy makers, funding agencies, other intermediaries, as well as research organizations. This includes a self-defined code of conduct and corresponding compliance rules of the science system as well.

A shift away from output or performance orientations to a process orientation is a mandatory prerequisite. Policy making therefore needs to change as well and develop such a process-perspective in terms of funding and monitoring. This will also overcome the challenges arising from causality and timeliness of impacts that do not fit with policymaking cycles or timelines. The policies should pay attention to the processes and their contributions to potential impacts, instead of the impacts themselves, as these are usually in the distant future when an impact on the impact generation process cannot be expected anyway. Policymaking needs to detach itself from the idea of an attribution of individual policies, programs or actions, but rather focus on contributions.

All together, policies as well as scientific contributions to solving societal challenges should be assessed against the background of their (potential) contribution to the generation of (future) impact. The framework to assess these potential contributions and impacts needs to be set normatively in societal negotiations, including all actors, or by normative rules of the science system (e.g. responsible research and innovation frameworks).

In any case, the dimensions against which science is assessed must be defined beforehand and cannot be defined ex-ante. The direction of the process can only be set at the beginning and not at the end of the particular process! For policy making as well as the self-organization of the science system, this means also that the settings and definition processes need to be organized and orchestrated (for example through public participation processes).

Finally, this also leads to a need of a shift of perspectives within the science system away from 'pure verocraty' to the acceptance of normativity in the agenda setting. 'Dare more normativity!' was the repeated plead in one of the workshops.

7 Summary and outlook

The literature survey as well as two workshops with national and international experts revealed a clear demand, even a need, to pay more attention to the societal impact of science and science systems – more than it has been done in the past in Germany. We understand impact as a contribution that science/research makes to society. This implies that society – or its sub-systems – defines the directions and goals of these contributions. Societal challenges or societal development goals might provide a general framework for such definitions.

With the expectation of the generation of impact by science actors, three main challenges directly occur. The first one is the attribution problem, namely the fact that impacts usually arise from multiple pathways and inputs so that individual causal-effect-relationships are difficult/impossible to identify and sometimes may not even make sense to define. The second challenge is adequate measurement and assessment, which not only lacks relevant data and indicators, but also an understanding and modelling of the dynamics of the systems. Third, the time delay of impact occurrence in relation to input, output or outcomes is related to the first challenge of attribution, but also leads to the necessity of a conceptual detachment of policy interventions and impact generations that is beyond usual policy cycle timelines.

It is not sufficient to add simply another column to the already too simplistic linear model of science and research production that leads from input via output to outcome and finally impact. In particular, the expectation of measuring the impact by traditional or even newly elaborated indicators is too shortsighted due to the multi-causality and delayed occurrence of most impacts.

When setting the impact goals, it has to be acknowledged that different types of impacts – particularly mentioned were economic, scientific/technological, environmental, policy, cultural, or health impacts – might be in contradiction to or result in a trade-off of one impact type versus the other.

Societal challenges as a guiding framework for policymaking are meanwhile widely accepted. A systemic perspective will help address these challenges, from our point of view. In order to allow science and research contribute to addressing these challenges and to generate societal impact, a fundamental change in science policy making and especially science evaluation is necessary. However, there is still an implementation challenge for such policies that manifests itself in different regards. Given that single causalities and especially direct and timely impacts cannot be expected, defining aspired impacts as well as setting impact goals are inevitable, so that the science system or the scientific organizations can do their best to achieve these agreed aims. Their efforts to achieve these aims is what should be evaluated and assessed, instead of the generated impact itself. Furthermore, the definition of the impact goals for particular research performing organizations need to be mission-specific, taking into account the governance structures as well as disciplinary differences.

In reality, collaboration between a number of actors and stakeholders will occur, with varied competences and interests. Considering this heterogeneity in policymaking and policy evaluation, new/different forms of policy engagement are required. With a systemic perspective and overarching departmental or institutional responsibilities, joint and collaborative efforts are to be established, where all stakeholders – policymaking, administration, intermediaries as well as science organizations – should be seen as responsible in the provision of pathways for societal impact. This will result in an increased relevance of cross-departmental topic and goal definitions. In addition, questions of how to accompany missions, instead of just initiating or funding them will gain importance, which requires a different strategic positioning of politics and policies. A systematic perspective on impact is required to ensure that the impact processes are as effective as possible.

Conclusively, in order to be able to measure and assess the impact of science at the systemic level as well, new theoretical and empirical approaches need to be developed. If science and innovation policy is also to design policy interventions that increase impact, a different perspective from the current one must be adopted. At the same time, it is essential to broaden the conceptual and methodological basis for understanding and measuring the impact of the science system. Relevant stakeholders in the German science system in general, and the Federal Ministry of Education and Research in particular, should establish an expanded foundation for how the science system and its contributions are understood and assessed in the face of global challenges and increasing importance, and to fulfill the "new social contract for research". Resorting to old concepts and metrics, however important they remain in detail, is no longer sufficient.

Policy recommendations:

- A different policy style as well as different expectations should be established with regard to societal impact generation (because they are not directly measurable).
- There is a need to understand impact as a process rather than impact as an outcome.
- In consequence, stakeholders should setup a framework that aims to promote engaging in impact generation (productive interactions) rather than to focus and measuring impact itself.
- However, impact aims should be formulated beforehand and the science system or the scientific organizations should be evaluated with regard to the engagement in achieving these aims. Dare more normativity of these impact aims.

- The definition of these aims should be based on particular societal challenges and needs and should be defined based on joint understanding and agreement. Participation/representation of stakeholders from society, policy and science should be pursued.
- The difficulty to measure cause-and-effect relationships in impact generation renders the distribution of research funds based on past impact performance – however measured – problematic. Instead, impact measurement understood as a process should focus on facilitating learning by the involved actors.
- This can only be achieved by a radical shift in the evaluation philosophy towards learning-based approaches.

References

Aghion, P.; Howitt, P. (1990): A model of growth through creative destruction.

Allan, G.; Figus, G.; Schubert, T. (2022): Understanding the macroeconomic Effects of public Research: an Application of a Regression-microfounded CGE-model to the Case of the Fraunhofer-Gesellschaft in Germany.

Bence, V.; Oppenheim, C. (2005): The Evolution of the UK's Research Assessment Exercise: Publications, Performance and Perceptions. Journal of Educational Administration and History 37(2), 137–155.

Bilsen, V.; Isabelle De Voldere, I.; Van Hoed, M.; Zeqo, K. (2018): Economic Footprint of 9 European RTOs in 2015-2016, Final report, prepared for EARTO – European Association of Research and Technology Organisations, Brussels: EARTO.

Blume, L.; Fromm, O. (1999): Regionale Ausgabeneffekte von Hochschulen. Raumforschung und Raumordnung| Spatial Research and Planning, 57(5-6), 418-431.

BMBF (2014): Die neue Hightech-Strategie: Innovationen für Deutschland. Berlin: Bundesministerium für Bildung und Forschung.

BMBF (2017): Fortschritt durch Forschung und Innovation: Bericht zur Umsetzung der Hightech-Strategie. Berlin: Bundesministerium für Bildung und Forschung.

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.

Bos, M.W.; Hofman, E.; Kuhlmann, S.; (2016): An assessment method for system innovation and transition (AMSIT). International Journal of Foresight and Innovation Policy 11(4), 185.

Böttcher, W.; Kerlen, C.; Maats, P.; Schwab, O.; Sheikh, S. (Eds.) (2014): Evaluation in Deutschland und Österreich: Stand und Entwicklungsperspektiven in den Arbeitsfeldern der DeGEval - Gesellschaft für Evaluation. Waxmann, Münster, Westf, 224 S.

Bozeman, B. (2007): Public values and public interest: Counterbalancing economic individualism. Washington, D.C.: Georgetown University Press, https://www.jstor.org/stable/j.ctt2tt37c

Bozeman, B.; Sarewitz, D. (2011): Public Value Mapping and Science Policy Evaluation. Minerva 49(1), 1–23.

Bruno, N.; Kadunc, M. (2018): Commission proposal for the Next EU Research & Innovation Programme (2021-2027). Presentation on Research Working Party Horizon Europe, Brussels: European Commission.

Bruno, N.; Kadunc, M. (2019): Impact Pathways: Tracking and Communicating the Impact of the European Framework Programme for Research and Innovation, in fteval Journal for Research and Technology Policy Evaluation, 47, 62-71, DOI: 10.22163/fteval.2019.330.

Buxton, M. (2011): The payback of 'Payback': Challenges in assessing research impact. Research Evaluation 20(3), 259–260.

Buxton, M.; Hanney, S. (1996): How Can Payback from Health Services Research Be Assessed? Journal of Health Services Research & Policy 1(1), 35–43.

Carlsson, B.; Jacobsson, S.; Holmén, M.; Rickne, A. (2002): Innovation systems: Analytical and methodological issues. Research Policy 31(2), 233–245.

Carree, M.; Della Malva, A.; Santarelli, E. (2014): The contribution of universities to growth: Empirical evidence for Italy. The Journal of Technology Transfer, 39(3), 393-414.

Comin, D. (2021): A General Equilibrium Quantification of the Impact of Fraunhofer on the German Economy. München: Fraunhofer-Gesellschaft. https://www.fraunhofer.de/de/forschung/leistungsangebot/wirkung-von-fraunhofer-forschung.html

Comin, D.; Giarda, M. (2019): Final Report: Baseline model for FRAME. A Framework to Study the Macroeconomic Effects of Innovation Policy. https://ftp.zew.de/pub/zew-docs/veranstaltungen/FRAME/Deliverables/D1.2.pdf

Comin, D.; Licht, G.; Pellens, M.; Schubert, T. (2019): Do companies benefit from public research organizations? The impact of the Fraunhofer Society in Germany. The Impact of the Fraunhofer Society in Germany, 19-006. München: Fraunhofer-Gesellschaft. https://www.fraunhofer.de/content/dam/zv/de/leistungsangebot/wirkung-von-forschung/the-impact-of-fraunhofer.pdf

CSIRO (2015): Impact Evaluation Guide. Black Mountain ACT 2601, Australia: CSIRO, 48 pp. https://www.csiro.au/~/media/About/Files/Our-impact-framework/CSIROImpactEvaluationGuide_Nov2015_WEB.pdf?la=en&hash=B351D24FB3CE02CB34FB859F2C34AA3940EE6D1F (Accessed 22 February 2018).

Donovan, C. (2008): The Australian Research Quality Framework: A live experiment in capturing the social, economic, environmental, and cultural returns of publicly funded research. New Directions for Evaluation 2008 (118), 47–60.

Donovan, C.; Hanney, S. (2011): The 'Payback Framework' explained. Research Evaluation 20(3), 181–183.

Engels, D. (Ed.) (2013): Anforderungen an Wirtschaftlichkeitsuntersuchungen finanzwirksamer Maßnahmen nach: [Empfehlungen des Präsidenten des Bundesrechnungshofes als Bundesbeauftragter für Wirschaftlichkeit in der Verwaltung]. Stuttgart: Kohlhammer, 149 S.

ERiC (2010): Evaluating the societal relevance of academic research: A Guide 1001 EN. Evaluating research in Context, 24 pp. http://www.siampi.eu/Content/ERiC%20Guide%202010.pdf

Federal Republic of Germany (2017): § 7 Bundeshaushaltsordnung: BHO.

Feidenheimer, A.; Frietsch, R.; Schubert, T.; Neuhäusler, P. (2019): Final report on the conceptual framework & proposed indicators. European Commission. https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5d1e20d9f&appId=PPGMS

Feidenheimer, A.; Frietsch, R.; Schubert, T.; Neuhäusler, P. (2018): Intermediate report on the conceptual framework, project report: Big Data approaches for improved monitoring of research and innovation performance and assessment of the societal impact in the Health, Demographic Change and Wellbeing Societal Challenge, Karlsruhe: Fraunhofer ISI; https://cordis.europa.eu/project/id/770531/results

Feller, I.; Stern, P.C. (2007): A strategy for assessing science: Behavioral and social research on aging. Washington (DC): National Academies Press, https://pubmed.ncbi.nlm.nih.gov/20669474/

Frietsch, R.; Lutz, J.; Neuhäusler, P.; Schubert, T.; Lerch, C.; Bethke, N.; Rothengatter, O. (2017): Der Beitrag der Fraunhofer-Gesellschaft zum Deutschen Innovationssystem. Endbericht zum Projekt, Karlsruhe: Fraunhofer ISI.

Frietsch, R.; Rammer, C.; Schubert, T.; Som, O.; Beise-Zee, M.; Spielkamp, A. (2015): Innovation Indicator 2015, acatech and BDI (Eds.), Berlin: acatech/BDI.

Frietsch, R.; Schmoch, U. (2006): Technological Structures and Performance as Reflected by Patent Indicators, In Schmoch, U.; Rammer, C.; Legler, H. (Eds.): National Systems of Innovation in Comparison. Structure and Performance Indicators for Knowledge Societies, Berlin: Springer.

Glorius, B.; Schultz, A. (2002): Die Martin-Luther-Universität als regionaler Wirtschaftsfaktor (Nr. 1). Halle.

Glückler, J.; Panitz, R.; Wuttke, C. (2015): Die wirtschaftliche Wirkung der Universitäten im Land Baden-Württemberg. Raumforschung und Raumordnung| Spatial Research and Planning, 73(5), 327-342.

Hare, P.G. (2003): The United Kingdom's Research Assessment Exercise. Higher Education Management and Policy 15(2), 43–62.

Hill, S. (2016): Assessing (for) impact: Future assessment of the societal impact of research. Palgrave Communications 2(1), 5.

Holmberg, K.; Bowman, S.; Bowman, T.; Didegah, F.; Kortelainen, T. (2019): What Is Societal Impact and Where Do Altmetrics Fit into the Equation? Journal of Altmetrics 2(1).

Jansen, D.; Wald, A.; Franke, K.; Schmoch, U.; & Schubert, T. (2007): Drittmittel als Performanzindikator der wissenschaftlichen Forschung. Kolner Zeitschrift fur Soziologie und Sozialpsychologie, 59(1), 125.

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. Research Evaluation, 24(4), 440-453.

Joly, P.-B.; Matt, M. (2017): Towards a new generation of research impact assessment approaches. The Journal of Technology Transfer 1 (4), 537.

Kaiser, U.; Kuhn, J. M. (2012): Long-run effects of public–private research joint ventures: The case of the Danish Innovation Consortia support scheme. Research Policy, 41(5), 913-927.

KNAW, VSNU, NWO (2014): Standard Evaluation Protocol 2015-2021 (SEP): Protocol for Research Assessments in the Netherlands (Amended version, Sep. 2016), 32 pp.

Kuhlmann, S.; Arnold, E. (2001): RCN in the Norwegian research and innovation system. Technopolis Group.

Martin, B.R. (2011): The Research Excellence Framework and the 'impact agenda': Are we creating a Frankenstein monster? Research Evaluation 20 (3), 247–254.

Matt, M.; Gaunand, A.; Joly, P.-B.; Colinet, L. (2017): Opening the black box of impact – Ideal-type impact pathways in a public agricultural research organization. Research Policy 46(1), 207–218.

Molas-Gallart, J. (2012): Research governance and the role of evaluation: A comparative study. American Journal of Evaluation 33(4), 577–592.

Molas-Gallart, Jordi (2015): Research evaluation and the assessment of public value. Arts & Humanities in Higher Education 2015, 14(1) 111–126.

Morgan, B. (2014): Research impact: Income for outcome. Nature 511(7510), 72-75.

Nowotny, H.; Scott, P.; Gibbons, M. (2003): Introduction: 'Mode 2' Revisited: The New Production of Knowledge. Minerva 41(3), 179–194.

Nowotny, H.; Scott, P.; Gibbons, M. (2008): The new production of knowledge: The dynamics of science and research in contemporary societies, Reprinted. ed. Sage Publ, London, IX, 179 Seiten.

OECD, Eurostat (2005): Oslo Manual: Guidelines for Collecting and Interpreting Innovation Data, 3rd Edition, 3rd edition ed. OECD Publishing, Paris, 1 online resource (162.

Raftery, J.; Hanney, S.; Greenhalgh, T.; Glover, M.; Blatch-Jones, A. (2016): Models and applications for measuring the impact of health research: Update of a systematic review for the Health Technology Assessment programme. Health technology assessment (Winchester, England) 20(76), 1–254.

REF (2012): Assessment framework and guidance on submissions. HEFCE, 63 pp. http://www.ref.ac.uk/2014/media/ref/content/pub/panelcriteriaandworkingmethods/01_12.pdf.

REF (2014): Research Excellence Framework 2014: The results. HEFCE. http://www.ref.ac.uk/2014/media/ref/content/pub/REF%2001%202014%20-%20full%20document.pdf

Robin, S.; Schubert, T. (2013): Cooperation with public research institutions and success in innovation: Evidence from France and Germany. Research Policy, 42(1), 149-166.

Romer, P.M. (1990): Endogenous technological change. Journal of political Economy, 98(5, Part 2), S71-S102.

Roy, G.; Allan, G.; Figus, G.; Knoche, A. (2021): The macroeconomic impact of Fraunhofer-Gesellschaft: A CGE approach, using micro-evidence Fraser of Allander Institute University of Strathclyde Business.

Schubert, T. (2021): The macroeconomic effects of the Fraunhofer-Gesellschaft.

Schmoch, U.; & Schubert, T. (2009): Sustainability of incentives for excellent research—The German case. Scientometrics, 81(1), 195-218.

Schubert, T. (2021): The macroeconomic effects of the Fraunhofer-Gesellschaft. https://www.fraunhofer.de/content/dam/zv/de/forschung/leistungsangebot/The-macroeconomic-effects-of-the-Fraunhofer-Gesellschaft.pdf

Schubert, T.; Baier, E.; Hufnagl, M.; Meyer, N;, Schricke, E.; Stahlecker, T. (2012): Endbericht zur Metastudie Wirtschaftsfaktor Hochschule. Stifterverband für die Deutsche Wissenschaft. Fraunhofer-Institut für System-und Innovationsforschung ISI. Karlsruhe.

Schubert, T.; Kroll, H. (2013): Endbericht zum Projekt „Hochschulen als regionaler Wirtschaftsfaktor". Karlsruhe. Online: http://www.stifterverband.de/wirtschaftsfaktorhochschule/regionale_bedeutung_von_hochschulen.pdf

Schubert, T.; Kroll, H. (2016): Universities' effects on regional GDP and unemployment: The case of Germany. Papers in Regional Science, 95(3), 467-489.

Spaapen, J.; Van Drooge, L. (2011): Introducing 'productive interactions' in social impact assessment. Research evaluation, 20(3), 211-218.

Spaapen, J.; van Drooge, L.; Propp, T.; van der Meulen, B.; Shinn, T.; Marcovich, A.; van den Besselaar, P.; Jong, S. de, Barker, K.; Cox, D.; Morrison, K.; Sveinsdottir, T.; Pearson, D.; D'Ippolito, B. (2011): SIAMPI final report. KNAW, 36 pp. http://www.siampi.eu/Content/SIAMPI_Final%20report.pdf.

SRC (2015): Research quality evaluation in Sweden - FOKUS: Report of a government commission regarding a model for resource allocation to universities and university colleges involving peer review of the quality and relevance of research. Swedish Research Council, Stockholm, 118 pp.

Sundbo, J. (1991): Strategic paradigms as a frame of explanation of innovations: A theoretical synthesis. Entrepreneurship & Regional Development, 3(2), 159–173.

Sundbo, J. (1995): Three paradigms in innovation theory. Science and Public Policy 22(6), 399–410.

Sundbo, J. (1998): The theory of innovation: Entrepreneurs, technology and strategy, Reprinted. 2003 ed. Elgar, Cheltenham, IX, 220 pp.

TEP, CEPS (2010): Study on Social Impact Assessment as a tool for mainstreaming social inclusion and social protection concerns in public policy in EU Member States. European Commission, 116 pp.

Wilsdon, J.; Allen, L.; Belfikore, E.; Campbell, P. (2015): The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management. DOI: 10.13140/RG.2.1.4929.1363

Glossary

BHO	Bundeshaushaltsordnung (Federal Budget Regulation)
BMBF	Bundesministerium für Bildung und Forschung (Federal Ministry of Education and Research)
CSIRO	Commonwealth Scientific and Industrial Research Organisation
ERiC	Evaluating Research in Context
GG	Grundgesetz (Basic Law)
GWK	Gemeinsame Wissenschaftskonferenz (Joint Research Conference)
HE	Horizon Europe
HEFCE	Higher Education Funding Council for England
HTS	High Tech Strategy
PRO	Public Research Organisation
RAE	Research Assessment Exercise
R&D	Research and Development
REF	Research Excellence Framework
R&I	Research and Innovation
RPO	Research Performing Organisation
RQF	Research Quality Framework
SEP	Dutch Standard Evaluation Protocol
TBE / TBIE	Theory-based evaluation / Theory-based impact evaluation
ToC	Theory of Change
UK	United Kingdom