

# Evaluierung neu denken

Studie im Auftrag der Plattform für Forschungs-  
und Technologiepolitikevaluierung (FTEval)

ENDBERICHT

September 2020

WWTF-G 2019/09



Autoren

Michael Stampfer & Michael Strassnig (WWTF GmbH)

WWTF GmbH

Schlickgasse 3/12

1090 Wien

# Inhaltsverzeichnis

0 Executive Summary	3
English summary and key recommendations	3
Kurzzusammenfassung und Empfehlungen	5
1 Auftrag und Studiendesign	9
2 Zentrale Rahmenbedingungen und Kontexte für FTI-Politik-Evaluierung in Österreich	11
Das FTI-Politik Umfeld – Status quo	11
Das FTI-Politik Umfeld – Aktuelle Entwicklungen	13
Marktsituation und Akteure	16
Akademisches Umfeld	19
Initiativen aus anderen Ländern, die uns etwas sagen können	21
Datenzugang und Datenverfügbarkeit	30
3 Befunde aus den Interviews mit Stakeholdern und Expert*innen	36
Zur Befragung	36
Generelle Reflexionen über die relevante Umwelt der FTI-Politik Evaluierung	36
Neue Entwicklungen und Herausforderungen für FTI-Politik-Evaluierung	38
Relevanz unterschiedlicher Daten für die FTI-Politik-Evaluierung	40
Methoden in der FTI-Politik-Evaluierung	44
Rahmenbedingungen bei beauftragten Evaluationen	49
Humanressourcen und Rekrutierung	51
Sichtweisen auf öffentlich finanzierte Supportmaßnahmen	52
4 Empfehlungen	56
Argumentation einer staatlichen Interventionslogik	57
Datenzugang, Datenzugang, Datenzugang!	58
Vorschläge für Maßnahmen auf Auftraggeberseite	59
Vorschläge für öffentliche Supportmaßnahmen	60
5 Literatur	63
6 Anmerkungen	65

## Abkürzungsverzeichnis

AI	Artificial Intelligence, Artificial Intelligence
AIT	Austrian Institute of Technology
AMDC	Austrian Micro Data Center
aws	Austria Wirtschaftsservice, Austria Wirtschaftsservice
BMBF	Bundesministerium für Bildung und Forschung (Deutschland)
BMBWF	Bundesministerium für Bildung, Wissenschaft und Forschung (Österreich)
BMDW	Bundesministerium für Digitalisierung und Wirtschaftsstandort (Österreich)
BMF	Bundesministerium für Finanzen (Österreich)
BMK	Bundesministerium für Klimaschutz, Umwelt, Energie, Mobilität, Innovation und Technologie (Österreich)
BWL	Betriebswirtschaftslehre
CDG	Christian Doppler Gesellschaft
CWTS	Centre for Science and Technology Studies, Leiden University
DSGVO	Datenschutzgrundverordnung
DUK	Donauuniversität Krems
EB	Erläuternde Bemerkungen
ELGA	Elektronische Gesundheitsakte
EQAR	European Quality Assurance Register for Higher Education
FFG	Österreichische Forschungsförderungsgesellschaft
FH OÖ	Fachhochschule Oberösterreich
FOG	Forschungsorganisationsgesetz
FTB	Österreichischer Forschungs- und Technologiebericht
FTI	Forschung, Technologie und Innovation
FWF	Fonds zur Förderung der wissenschaftlichen Forschung (Österreich)
GSK	Geistes-, Sozial- und Kulturwissenschaften
IFZ	Interuniversitäres Forschungszentrum für Technik, Arbeit und Kultur
ITF	Innovations- und Technologiefonds
ML	Machine Learning
NSF	National Science Foundation (U.S.)
NWO	Nederlandse Organisatie voor Wetenschappelijk Onderzoek
ÖAW	Österreichische Akademie der Wissenschaften
OeAD	Österreichischer Austauschdienst
OECD	Organisation for Economic Co-operation and Development
Plattform FTEval	Österreichische Plattform für Forschungs- und Technologiepolitikevaluierung
PRIME	Policies for Research and Innovation in the Move towards the European Research Area
RCN	Research Council of Norway
RFT	Rat für Forschung und Technologieentwicklung
RIS	Research Information System
RISIS	Research Infrastructure for Science and Innovation Policy Studies
RoRI	Research on Research Institute
RRI	Responsible Research and Innovation
RTI	Research, Technology and Innovation
SciSIP	Science of Science and Innovation Policy Program
SDG	Societal Development Goals
SIPER	Science and Innovation Policy Evaluations Repository
SNSF	Swiss National Science Foundation
SPRU	Science Policy Research Unit
STS	Science and Technology Studies
TEKES	Finnish Funding Agency for Technology and Innovation
TIPC	Transformative Innovation Policy Consortium
ToR	Terms of Reference
USP	Unique Selling Position
VWL	Volkswirtschaftslehre
WU Wien	Wirtschaftsuniversität Wien

# 0 Executive Summary

## *English summary and key recommendations*

This study examines concepts for public support actions to strengthen the methodological capabilities of market actors in the area of RTI policy evaluation against the backdrop of current and upcoming challenges for this area, in particular the increasing complexity of RTI politics in many countries. Existing approaches, methods, indicators and data sets run at risk to reach their limits and to fail to meet the market demand.

The market for RTI-policy evaluation in Austria is characterised by a comparatively (very) small size both in terms of number and budget of contracts as well as by a comparably high number of national and international providers of evaluations compared to the small market size. Evaluations of individual RTI programs are frequently commissioned as this type of intervention dominates Austrian RTI policy. Rare is the number of exercises that target institutions, specific thematic areas or the RTI system as such. Contractors mainly come from the private and (semi-)public extra-mural social sciences sector. This sector has developed a USP regarding several RTI-related approaches and themes over the last decades, with a near complete absence of higher education institutions in that area. While there is a constant buzz about governance, strategies and missions, policy measures are being dominantly structured around sector-specific or collaborative RTI programmes “owned” by particular ministries and agencies, with implications for evaluations resulting in formative, i.e. learning or in legitimizing approaches on the program level as well as the tool-case of established evaluation methods. Existing administrative and statistics RTI data as a rule cannot be used due to legal regulations and particular interests.

International examples show that also broader approaches are possible and being followed both by contractors and academics / evaluators: In other countries we can find larger evaluations with more ambitious concepts to grasp broader policy issues and larger policy interventions, e.g. around mission-oriented policy approaches and instruments. Further, we can identify serious and structured work to capture new realities with new or improved indicators and methods.

Evidence for this study has been collected mainly via desktop research and about 20 interviews mainly with national members of the RTI policy evaluation community and some international experts. With the former, semi-structured interviews on general conditions, methods and data have been conducted. The interviews stressed the relevance of data access and a changed RTI policy allowing for methodological innovations. Within the existing framework dominated by evaluations on programme level, descriptive statistics and qualitative methods (interviews, survey) are regarded as bread and butter methods – also in the foreseeable future. Potential benefits of more recent methods (e.g. from data science) are not yet clearly visible to the interviewees. The interviewees are also dominantly sceptical regarding a targeted public support action specific to RDI-policy evaluation providers.

### *Recommendations regarding data access*

- Focus on the implementation of the planned *Austrian Micro Data Centers AMDC* to open up and better link data repositories of *Statistik Austria* and other public registers. AMDC allows for access in line with GDPR requirements for personal data.
- Open up registers and administrative data sets, including workable links to the AMDC; further install the long-awaited *Forschungsförderungsdatenbank* (research funding data bank; FFDB) and consider a common data repository hosted by FTEVal.

### *Recommendations for contracting bodies*

- Act strategically; this includes the option of granting of a higher degree of autonomy to funding agencies. The agencies then can act with broader funding portfolios while ministries can focus on policy portfolios. In the set up of the new *Forschungsfinanzierungsgesetz* (FoFinaG, research finance act), there are clear opportunities in the upcoming months to (i) make a next step in programme design and ownership as well as to (ii) better provide schedules and resources for proper evaluations.
- Grant larger evaluations; within the format of funding programmes a reduction in numbers of programmes can lead to less but bigger initiatives which then can lead to larger evaluations with a richer array of methods and approaches. A focus on missions could shift evaluations from legitimization more to learning and impacts.
- Larger contracts also allow the evaluators to (1) experiment with a broader set of methods within their assignment and to (2) earn money and re-invest part of it into method development, further qualification and other activities.

### *Recommendations for public support measures*

This study does not come forward with a specific narrow support programme for evaluators to get method development funded. However, three other measures might help strengthen social sciences and RTI policy research both in the non-university and university realm.

- More than ten years ago, the science ministry had offered a programme called *Dynamische Qualitätssicherung* (DQS, dynamic quality assurance) to support those non-academic social science research actors with low or no block grants. This programme should be reanimated in the light of the COVID-19 crisis to support continuing education, method development, interdisciplinary work, preparation of larger projects and other activities to strengthen the sector.
- A larger social science funding initiative for applied but high quality (empirical) research might also help to support research – hand in hand with the opening of public data repositories.
- Finally, capacity building should take place at the Austrian universities, e.g. with endowed professorships in the field of RTI policy including method development. These professorships should come as double affiliations with non-university research institutes to safeguard collaboration and care for critical mass.

## **Kurzzusammenfassung und Empfehlungen**

Diese Studie geht der Frage nach, in welcher Form öffentliche Unterstützungsmaßnahmen getroffen werden können, um im Feld der FTI-Politik-Evaluierung in Österreich die Kapazitäten für die Entwicklung, Aneignung und Anwendung von avancierten Methoden zu erhöhen. Den Hintergrund dafür bilden zunehmend komplexe Konzepte und Interventionsformen in vielen FTI-Systemen. Daher drohen bestehende Routinen, Methoden, Indikatoren- und Datensätze sowie Auftragsformen für Evaluierungen ihre Grenzen zu erreichen und die Konkurrenzfähigkeit von Akteuren jedenfalls auf internationalen Märkten einzuschränken.

Der österreichische Markt für FTI-Evaluierungen ist sehr klein, von kleinstrukturierten Aufträgen ebenso geprägt wie von einer wachsenden Zahl von Anbieter\*innen aus dem Inland und teils auch aus dem Ausland. Das Hauptprodukt sind Evaluierungen zu einzelnen FTI-Förderprogrammen. Seltener werden Arbeiten beauftragt, die auf Institutionen, größere thematische Felder oder das FTI-System insgesamt abzielen. Die Anbieter\*innen kommen überwiegend aus dem Feld der privatwirtschaftlichen oder (halb-)öffentlichen außeruniversitären Sozialwissenschaften. Dieser Community ist es auch gelungen, einen USP zu erlangen und diesen teils auch in EU- und andere internationale Aufträge umzumünzen. Der akademische Sektor hingegen ist sehr schwach ausgeprägt und stellt weder für FTI-Politikgestaltung noch für ihre Evaluierung nennenswerte Kapazitäten bereit.

Das erwähnte „Hauptprodukt“ Programmevaluierung steht auch deshalb sehr stark im Vordergrund, weil in der heimischen FTI-Politik das Instrument des abgegrenzten, lang fortgeführten Förderprogramms dominiert, das in der Ownership von einzelnen Ministerien und / oder Agenturen steht. An diesem Instrument wird gelernt und Legitimierung gesucht und es beeinflusst die gewählten Pfade und Methoden bei Evaluierungen. Die Legitimierung erfolgt überwiegend gegenüber den relevanten Mitspieler\*innen im Policy-System und auf einer Datenbasis, die von schweren Restriktionen geprägt ist, soweit es sich um staatlich erhobene (Register- oder Mikro-)Daten handelt.

Hier setzt auch die Analyse an: Was steht für die österreichischen Policy-Akteure im Vordergrund, wenn sie tätig sind, und welche Werkzeuge schaffen sie sich bzw. bekommen sie in die Hand gegeben? Studien und Evaluierungen (von der Systemevaluierung bis zum OECD Review) beobachten eine Vielfalt von Interventionsformen, freilich auf einer schmalen Bandbreite. Diese schmale Bandbreite und die zahlreichen kleineren Programme determinieren dann die Art und Methodenwahl der Evaluierungen. Daher scheint es auch keine bedeutende Diskrepanz zwischen Angebot und Nachfrage zu geben.

Dem steht ein stärker werdender internationaler Trend gegenüber, der in Richtung Missionen, breiteren Instrumenten, stärkerer Datenorientierung und stärkerer gesellschaftlicher Legitimation geht. Das bedingt und ermöglicht andere Formen der Evaluierung, was sich auch in den untersuchten internationalen Beispielen zeigt: Andere Methoden, andere und umfangreichere Evaluierungen sind möglich, weil ein anderes Policy-Verständnis möglich ist, nämlich größere, zusammenhängende Probleme angehen zu wollen und dazu auch Sektor- und Organisationsgrenzen besser überwinden zu wollen. Dieser internationale Trend kommt in Österreich nur zögerlich und noch kaum auf der Instrumentenebene an.

Für Österreich bietet nun die im Juli 2020 in Kraft gesetzte Forschungsfinanzierungsnovelle, namentlich das Forschungsfinanzierungsgesetz (FoFinaG) eine vielversprechende Möglichkeit, einen nächsten Schritt in Richtung Programmgestaltung und Evaluierung zu

setzen. In den nächsten Monaten wird entlang der Stationen *Strategie – FTI-Pakt – Finanzierungsvereinbarungen* auch zu verhandeln sein, was passieren soll, mit welchen Instrumenten die Ziele erreicht werden sollen, ob die *principal-* oder die *agent-*Ebene die Programm-Ownership bzw. die über die Evaluierungen haben werden. Wie auch immer das Ergebnis im Einzelnen aussehen wird: Ein klarerer Interventions- und Evaluierungsrahmen kann auf verschiedene Weisen hergestellt werden; wichtig ist, dass er hergestellt wird.

Die Evidenz für diese Studie wurde überwiegend durch Desktop-Recherche und gut 20 Interviews zusammengetragen. Die Interviews erfolgten hauptsächlich mit Vertreter\*innen österreichischer Evaluierungsanbieter\*innen und -nachfrager\*innen, mit Datenexpert\*innen sowie mit internationalen Auskunftspersonen. Die Gespräche mit den österreichischen Interviewpartner\*innen erfolgten entlang semi-strukturierter Fragebögen, die auf Rahmenbedingungen, Methoden- und Datenfragen eingingen. Die Interviews zeigen die große Bedeutung eines besseren Datenzugangs ebenso auf wie die Notwendigkeit anderer Policy-Zugänge, um methodische Innovationen zu ermöglichen bzw. zum Einsatz zu bringen. Innerhalb der dominanten Form der Einzelprogramm-Evaluierungen werden deskriptive Statistik und qualitative Methoden wie Interviews und Surveys als das Grundgerüst angesehen und auch für die nächsten Jahre erwartet. Das Potenzial neuerer Methoden etwa aus dem Bereich der Data Sciences ist für viele der Befragten hingegen noch nicht offensichtlich.

Schließlich besteht eine weitgehende Skepsis bei den Befragten, was ein sehr spezifisches, maßgeschneidertes Methodenentwicklungs-Förderprogramm für die Community betrifft. Diese Studie liefert somit auch kein Programmdesign – insbesondere vor dem Hintergrund, dass (i) objektive Gründe (Marktgröße, Teilnehmer\*innenzahl, Transferierbarkeit von Ergebnissen, langfristige Wirkungen, etc.) dagegen sprechen sowie (ii) die Bandbreite der Haltungen der Interviewpartner sehr groß und über weite Strecken eben von Skepsis geprägt ist.

#### *Empfehlungen zum Datenzugang*

- Implementierung des *Austrian Micro Data Centers AMDC*, mit welchem die Datenbestände der Statistik Austria sowie weiterer Verwaltungsregister zugänglich gemacht und untereinander verknüpfbar werden können. Das AMDC bietet eine datenschutzkonforme Zugangsmöglichkeit auch zu personenbezogenen Daten.
- *Register und Verwaltungsdatensätze frei geben*. Des Weiteren wird den an FTEval beteiligten Ministerien stark empfohlen, dass Register und Verwaltungsdaten in ihrem Wirkungsbereich in das AMDC eingespeist werden und so auch für FTI-Politik-Evaluierung auf einfache Weise und einheitlich genutzt werden können. Im Einzelnen wird auch die *Einrichtung der (bereits angedachten) Forschungsförderungsdatenbank (FFDB)* empfohlen. Zusätzlich gibt es auch Möglichkeiten, dass die FTEval-Community im eigenen Bereich um eine verbesserte Datenlage sorgt: durch die Errichtung eines *gemeinsamen Datenrepositoriums* (im FTEval Repository).

#### *Empfehlungen für Maßnahmen auf Auftraggeber\*innenseite*

- *Strategisch agieren*: Dazu gehört auch, dass ein Umsetzen der Ergebnisse der aws / FFG Evaluierung die Verantwortung für die Körnigkeit und Taktung von Programmen von den Ressorts stärker hin zu den großen Förderagenturen verlagern kann. Dies findet seine Ergänzung darin, dass sich die Ministerien stärker um mis-

sionsorientierte Policy-Portfolios kümmern können. Für solche Fragen gibt es mehrere Optionen im Rahmen der neuen FTI-Strategie und der FoFinaG-Konkretisierung im Rahmen von FTI-Pakt und Finanzierungsvereinbarungen.

- *Größere Aufträge (1)*: Selbst wenn die traditionelle Programmförderung weiterhin das Instrument der Wahl bleibt, sollte es in Österreich weniger kleinstrukturierte Programme geben und einige von ihnen sollten auch ein Enddatum haben dürfen. Das würde erlauben, neben legitimierenden Zwischenevaluierungen viel mehr Raum und Ressourcen frei zu machen, um über andere Evaluierungen nachzudenken (und damit auch über neue Methoden) und diese auch durchzuführen sowie zu Vollkosten zu finanzieren.
- *Größere Aufträge (2)*: Autonomere Agenturen, die ihr Portfolio bespielen, können Evaluierungseinrichtungen ermöglichen, im Rahmen von großen Evaluierungen zu experimentieren, zu lernen und das Geld zu verdienen, das ein\*e Dienstleister\*in eben braucht, um seine Produkte auf hohem Niveau anbieten zu können.

#### *Empfehlungen für öffentliche Supportmaßnahmen*

Auch wenn aufgrund starker Argumente eine Supportmaßnahme spezifisch für die österreichische FTI-Politik-Evaluierungscommunity nicht empfohlen wird, gibt es in den Sozialwissenschaften insgesamt, gerade was eine angewandtere Ausrichtung betrifft, Wirkungsbereiche, die die Nöte dieses Bereichs lindern können und zu einer gesteigerten Qualitätsorientierung beitragen können. In diesem Rahmen können auch methodische Fragen einen Raum einnehmen.

- Wiederauflegen des Programms *Dynamische Qualitätssicherung DQS*, wie es bereits einmal vor weit mehr als zehn Jahren durchgeführt wurde. Das ist auch angesichts der Corona-Ausfälle ein Instrument, damit die Kleineren bzw. solche, die stark von Aufträgen abhängen, insgesamt überleben können. Im Programm wurde eine Vielzahl von Zielen verfolgt, es umfasste Maßnahmen wie Methodenerwerb, Weiterbildung, disziplinen- und institutionenübergreifendes Arbeiten, Anschubfinanzierung für Größeres, Publikationen, etc.
- *Größeres Forschungsprogramm für die Sozialwissenschaften im angewandten Bereich*, das Bereiche ins Visier nimmt, die aufgrund der knappen Mittel und der Ausrichtung beim FWF nicht finanziert werden können. Das heißt: Fokus auf Projektförderung für Fragen, die anwendungsnahe sind und das auch zeigen müssen; neue interdisziplinäre Ansätze und Methoden ausprobieren, wo ein akademischer Pay-off an Publikationsleistungen nicht unmittelbar gegeben ist; Nutzer\*innen des produzierten Wissens von Anfang an mit an Bord nehmen. Der gesellschaftliche Nutzen muss demonstriert werden anhand konkreter Objekte und Themen, die aus der Praxis kommen. Die Förderbedingungen müssten dergestalt sein, dass ein solches Programm auch für Institutionen attraktiv ist, die über keinerlei Basisfinanzierung verfügen.
- Ausbau der *akademischen Kapazitäten* in Bereichen, die für FTI-Politik-Evaluierung relevant sind. Es geht um Kapazitätsaufbau, Grundlagenarbeiten und Ausbildung einschließlich PhD, wie dies auch in praktisch allen anderen Ländern der Fall ist. Das Instrument der Wahl wäre das – um einige Elemente deutlich anzupassende – Instrument der Stiftungsprofessur. Professuren sollten so eingerichtet werden, dass Kooperationen mit dem außeruniversitären Bereich zu ihrer DNA



gehören, etwa über „Double Affiliations“: Damit sind auch größere Ressourcen durch Pooling möglich, und damit die Attraktion von besonders leistungsstarken Forscher\*innen. Es gilt bei der Ausgestaltung allerdings noch eine Vielzahl von Herausforderungen hinsichtlich Widmung, Auswahl und Finanzierung zu adressieren.

# 1 Auftrag und Studiendesign

Die WWTF GmbH erhielt den Auftrag, ein Konzept dafür zu entwickeln, wie strukturelle Unterstützungsmaßnahmen zum methodischen Kompetenzaufbau von Evaluierungsanbietern aussehen können. Hintergrund ist die Herausforderung, dass neue FTI-Evaluierungsherausforderungen bewältigt werden müssen: Die laufende Weiterentwicklung des FTI-Systems, steigende Systemkomplexität und den damit verbundenen methodisch-analytischen Herausforderungen sind dafür die wichtigsten Stichworte. Damit verbunden ist die These, dass das herkömmliche Set an Indikatoren, Daten und Methoden in der österreichischen Evaluierungspraxis an seine evaluativen Grenzen stößt. Diese Grenzen des Herkömmlichen sind das Aufkommen neuer thematischer Aspekte (z.B. Digitalisierung, open innovation, public sector innovation, neue Forschungsinfrastrukturen), neue regulative Herausforderungen (z.B. DSGVO, open data, Registerdaten sowie Daten der amtlichen Statistik) sowie neue Entwicklungen auf methodisch-analytischer Ebene (z.B. data science, big data, neue Indikatoren, Verknüpfung von Datenbeständen). Dabei wird von Anbietern und Auftraggeber\*innen das Risiko gesehen, mittelfristig den internationalen Anschluss, namentlich am Markt der EU-Aufträge zu verlieren.

Ziel der Studie ist es, ein Maßnahmenkonzept zum methodischen Kompetenzaufbau zur Bewältigung neuer FTI-Evaluierungsmaßnahmen vorzuschlagen um (1) den neuen Evaluierungsanforderungen aus der Systemumwelt zu entsprechen, (2) um sowohl im europäischen Wettbewerb kompetitiv zu bleiben als auch (3) Auftraggeber\*innen aus Österreich internationale state-of-the-art Herangehensweisen bieten zu können. Das Maßnahmenkonzept soll auch ermöglichen, unabhängig von Aufträgen – die nur kleine Ausschnitte der FTI-Politik abdecken, da sie weitgehend einer Programmlogik folgen – das Verstehen systemischer Zusammenhänge ermöglicht werden.

Der Auftrag umfasst folgende Arbeitspakete:

- Screening internationaler Beispiele / Ansätze. Dies wurde überwiegend mittels Desktop Recherche durchgeführt sowie mittels 4 Interviews mit Expert\*innen, die zu den internationalen Beispielen Auskunft geben konnten (Schweden, Deutschland, Niederlande).
- Statuspräsentation (Startworkshop) in einer von FTEval organisierten Veranstaltung. Dieser Workshop fand am 11. Dezember 2019 unter reger Beteiligung der FTEval Mitglieder statt. Die Diskussionen sind in die Erstellung des Leitfadens für die darauffolgende Interviewphase eingeflossen.
- Erhebungsphase durch eine schriftliche Befragung sowie Interviews (10-15). Insgesamt wurde mit 19 Vertreter\*innen von FTEval Mitgliedorganisationen gesprochen; wenige Befragungen davon fanden schriftlich statt, die meisten in längeren Interviews (1,5-2 Stunden pro Interview). Es wurde mit allen Mitgliedsorganisationen bzgl. eines Interviews Kontakt aufgenommen. Drei Mitgliedorganisationen haben festgehalten, dass sie zur gegenständlichen Studie wenig beitragen können. Mit zwei Vertreter\*innen von Mitgliedorganisationen konnten die Interviews wegen COVID-19 nicht durchgeführt werden. Zusätzlich zu den FTEval Mitgliedorganisationen wurden 5 externe Expert\*innen im (vorwiegend) akademischen Umfeld in

Wien durchgeführt (Data Science, Netzwerkanalyse, Artificial Intelligence, EU-Programm-Evaluierung, Volkswirtschaft). Aus Datenschutzgründen bzw. wegen der Zusage der Vertraulichkeit werden die Namen der Interviewpartner\*innen in dieser Studie nicht genannt.

- Die Erhebung erfolgte mittels eines weitgehend gleichen Fragebogens nach Art der Mitgliedorganisation (beauftragende Institution / durchführende Organisation). Der Fragebogen war weitgehend qualitativ orientiert, beinhaltete aber zu vielen Fragen auch eine (mehr oder weniger) quantitative Einschätzung. In den eher deliberativ orientierten Interviews konnte (auch aufgrund des jeweiligen Arbeitsfokus der interviewten Personen bzw. ihrer Rolle in der Gesamteinstitution) die quantitative Einschätzung nicht bei jedem Punkt durchgehend abgefragt werden. Daher werden in dieser Studie keine quantitativen Auswertungen präsentiert, sondern die Hauptstoßrichtungen in den Bewertungen durch die Interviewpartner\*innen beschrieben.
- Diskussion des vorgeschlagenen Maßnahmenpakets mit potenziellen Nutzer\*innengruppen in Form eines von FTEval organisierten Workshops. Dieser Workshop war für Ende März geplant und wurde aufgrund der Maßnahmen im Rahmen der COVID-19 Pandemie kurzfristig abgesagt. In welcher Form dieser Workshop in der Folge stattfinden kann, muss erst entschieden werden.
- Das Ergebnis der Recherchen und Interviews wird in diesem Bericht festgehalten und dient als Ausgangspunkt für die weitere Diskussion zur etwaigen Umsetzung von Maßnahmen. Diese Studie liefert kein fertiges Programmdesign – insbesondere vor dem Hintergrund, dass (i) wesentliche Parameter eines Programms nicht feststehen und ein politischer Prozess dazu erst begonnen werden muss, sowie (ii) die Bandbreite der Haltungen der Interviewpartner sehr groß ist.

Die Autoren des Berichts haben sich die Freiheit genommen, den breiten Kontext von FTI-Politik-Evaluierung ausführlicher anzuschauen, da dieses Umfeld ein wesentlicher Faktor für den aktuellen Status aber auch für die Sinnhaftigkeit und Effektivität von etwaigen Maßnahmen darstellt.

Wir danken allen, die mit uns im Rahmen dieses Auftrages ausführliche Gespräche geführt haben.

## 2 Zentrale Rahmenbedingungen und Kontexte für FTI-Politik-Evaluierung in Österreich

### Das FTI-Politik Umfeld – Status quo

Evaluierungen im Bereich der FTI-Politik sind in Österreich seit den späten 1990er Jahren ein Instrument, um Wirkungen von Politikhandeln zu überprüfen und Lernen sowie (Weiter-)Entwicklung zu ermöglichen. Dieses Instrument hat im Einsatz von der Anzahl her schrittweise aber stetig zugenommen (Dinges & Schmidmayer, 2010; Streicher, Polt, & Unger, 2020), nicht aber unbedingt von der Höhe der eingesetzten Mittel her.

Die Verbreitung über die verschiedenen Instrumente dieses Politikfeldes folgt bestimmten Mustern: Dort wo der Legitimationsdruck ein Stück höher ist und eine zeitliche Befristung bzw. Taktung der Maßnahmen den üblicheren Fall darstellt, dort finden wir mehr Evaluierungen. Das betrifft neben der stark in die akademische Lehre ragende Bereich der Akkreditierung oder regelhafter Weiterfinanzierung tertiärer Einrichtungen vor allem den Bereich der Forschungs- und Innovationsförderung.<sup>1</sup>

In der FTI-Förderung kommen zwei Elemente zusammen: Zum einen gibt es eben seit den späten 1990er Jahren eine sehr hohe Anzahl solcher, häufig vergleichsweise kleiner Instrumente, die als Programm ausgestaltet sind (Pichler, Stampfer, & Hofer, 2007; Schibany, Jörg, & Nones, 2005; Streicher, Polt, & Unger, 2020; OECD, 2018). Zum anderen ist es Usus, die meisten von ihnen auch einer periodischen Evaluierung zu unterziehen, was auch nicht zuletzt ein Verdienst der Plattform FTEval gewesen ist. Diesem Entwicklungsschritt war mit dem Innovations- und Technologiefonds ITF bereits ein mehrjähriger Lernprozess vorangegangen, sodass 1997 Erfahrungen und Forderungen – eben für Programmevaluierungen – festgeschrieben werden konnten (Ohler, Hutschenreiter, & Sturn, 1997).<sup>2</sup>

Für die frühen 1990er Jahre gilt weiters festzuhalten, dass es bereits ein echtes Methodenentwicklungsprogramm für die Evaluierungscommunity gegeben hat: Dieses hieß MONITOR / SPEAR und wurde im 2. Rahmenprogramm von der EU-Kommission ausgeschrieben. Das Programm bewirkte einen methodischen Push hinsichtlich Entwicklung und Verbreitung.<sup>3</sup>

Die österreichische Evaluierungspraxis umfasst sowohl ex ante- als auch ex post-Bewertungen, vor allem aber Evaluierungen während (interim) der – oft langen – Laufzeit der Programme und vergleichbarer Initiativen. Wir finden sie in Österreich namentlich im Bereich der Technologie- und Innovationsförderung mit wirtschaftlichem Bezug und nur in deutlich geringerer Ausprägung bei der Förderung wissenschaftlicher Forschung. Daher stehen einige spezifische Fragestellungen häufig im Vordergrund des österreichischen FTI-Evaluierungsgeschehens, darunter die nach der *Wirkung auf Unternehmen, ihre Technologieentwicklung, Produktivität und ihren Markterfolg* bzw. auf einer anderen Frageebene nach dem *Vorhandensein, der Verfügbarkeit und der Publizierbarkeit entsprechender Datensätze*. Da es auch viele Kooperationsprogramme gibt, treten Fragen nach *Kooperationsintensität und -wirkungen* hinzu.

Ebenso nimmt aufgrund der Zunahme missionsorientierter Förderungen der Wunsch zu, *Wirkungen auf Sektorpolitiken bzw. hinsichtlich der gesellschaftlichen Wirksamkeit geförderter Aktivitäten* zu erkennen und zu belegen. Für die Analyse solcher Programme existieren zum gegenwärtigen Zeitpunkt deutlich mehr Wünsche und Absichten als Methoden, Kausalketten und zugängliche Daten, was aufgrund der Vielschichtigkeit, ja häufig Komplexität der Problemstellungen, Akteurslandschaften und Interventionsformen wenig verwundert. Dem versucht die Plattform FTEval mit Projekten, Konferenzen und anderen Maßnahmen entgegenzuwirken.

Die derzeit noch geltende FTI-Strategie aus dem Jahr 2011 der österreichischen Bundesregierung (Bundeskanzleramt et al., 2011) gibt zwar große Ziele für diesen Politikbereich vor, bildet aber nur einen groben Rahmen. Soweit es zu dieser Strategie Kritik gegeben hat, hat sie sich neben der am Fehlen eines Budgetpfades (FTB, 2016, S. 90) auf das Fehlen konkreter Zielsetzungen bezogen, aus denen sich operationale Ziele und Maßnahmen definieren lassen, welche wiederum durch Evaluierungen überprüft werden können. Der letztlich klar verfehlt „Weg zum Innovation Leader“ konnte zwar anhand von Indikatoren verfolgt werden, siehe die jährlichen Berichte zur wissenschaftlichen und technologischen Leistungsfähigkeit Österreichs des RFTE (zuletzt Rat für Forschungs- und Technologieentwicklung, 2020, S. 62), Aufgaben und Beiträge bestehender und neuer Initiativen und Institutionen sind jedoch deutlich schwerer abzuleiten. Evaluierung kommt in der Strategie nur am Rand vor: Insgesamt wurde in diesem Dokument zwar häufig auf die Systemevaluierung aus dem Jahr 2008 Bezug genommen, aber in der FTI-Strategie selbst gibt es keine nennenswerten Bezüge zu Evaluierungen, ihre Wichtigkeit für die FTI-Politik und ihren konkreten Einsatz.

Der OECD Innovationsreview von 2018 (OECD, 2018) erwähnt Evaluierungen als Instrument der FTI-Politik häufig und stützt sich in seinen Erkenntnissen auch auf eine Reihe solcher Studien. Das Vorhandensein zahlreicher Evaluierungen und einer lebendigen Evaluierungskultur "*supported by the rather unique Platform FTEval*" (S. 46) wird positiv hervorgehoben. Kritisch sieht die OECD das Fehlen von Portfolioansätzen und das geringe Ausmaß an Politikkoordination (S. 46): "*More regular evaluations are needed of portfolios of major public support instruments, applying international best practice as concerns methodology and data access*" (S. 14). Kritik gibt es weiters zu deutlichen Schwachstellen bei (Mikro-)Datenverfügbarkeit, -einsatz und Evidenzorientierung, vor allem was die anwendungsorientierte Forschungsförderung betrifft (S. 19): "*Nevertheless, quantitative impact assessments that seek to identify causal effects are rare. This is mainly due to limited access to and tight restrictions on the use of firm-level and administrative data for evaluation purposes. Austria lags behind international best practice in this regard. This constraint could become even more important over time, as there is reason to assume that the political need for accountability as to the effects of public spending will increase*" (S. 46). Ein Punkt besonderen Interesses ist dabei der Umgang Österreichs mit Evidenz bei der indirekten Forschungsförderung. Dementsprechend gehen die Empfehlungen in Richtung Datenverfügbarkeit, Portfolio- und andere größere Evaluierungen und *accountability* dort, wo das richtig große Geld fließt.

Die letzten Forschungs- und Technologieberichte (FTBs) der Bundesregierung bieten meist Übersichten und Beschreibungen von durchgeführten Evaluierungen, nicht aber bzw. nur am Rande systematische Erwägungen zum Zusammenhang von Policy und Evaluierung. Der FTBs der letzten Jahre haben ein Kapitel „FTI-Evaluierungskultur und -praxis“:

Das Kapitel ist weitgehend eine Aufzählung und Beschreibung von Evaluierungs- und Begleitaktivitäten. In den sonstigen Teilen treten einzelne Evaluierungen als legitimatorische Zeugnisse auf. An systemischen Forderungen findet sich (nur): „Portfolio Evaluierungen durchführen; Regelmäßige Evaluierungen der Instrumentenportfolios sollen mit einem angemessenen Methodenmix durchgeführt werden und Zugriff auf die dahinter liegenden Daten erlauben.“ (FTB, 2019, S. 193). Der Forschungs- und Technologiebericht 2018 (FTB, 2018, S. 196ff.) beschäftigte sich ausführlicher mit Datenzugang zu Bundesstatistik und Registerdaten und der Beschreibung aktueller Evaluierungen.

Im Forschungs- und Technologiebericht 2017 (FTB, 2017) ging es neben einzelnen Evaluierungen (v.a. Forschungsprämie) auch um grundlegendere Fragen, da hier dezidiert Fragen der Evaluierungskultur in Österreich diskutiert wurden.<sup>4</sup> Ausgehend von positiven Grundbefunden (S. 146) werden einzelne kritische Punkte und Desiderate angesprochen: „... steigen die Anforderungen an die Methodik und Datenverfügbarkeit bzw. -analyse. Dabei gewinnt der Einsatz entsprechender Mess-Indikatoren, kontrafaktischer Analysen bzw. Kontrollgruppenvergleiche und quasi-experimenteller Methoden an Bedeutung.“ (S. 147f.) „Wenngleich sich Lücken im Bereich der Wirkungsevaluierung und im Zugang zu bzw. der Verknüpfung von statistischen Daten finden, wie etwa rezente Untersuchungen zeigen ..., sind die Aktivitäten und der Diskurs in der FTI-Politik vom Bemühen geprägt, die Evaluationskultur weiter zu stärken und das Programmdesign und Evaluierungsmethoden von Forschungs- und Technologieprogrammen zu verbessern. ... FTI-Evaluierungen müssen, dem internationalen Trend folgend, ihr Methodenset konzeptionell erweitern, um besser der Forderung von Wirkungsmessung entsprechen zu können. Dazu zählt beispielsweise der Einsatz von kontrafaktischen und quasi-experimentellen Methoden als Analyseinstrumente.“ (S. 148).

Der Forschungs- und Technologiebericht 2016 (FTB, 2016) schließlich enthält die üblichen Beschreibungen der durchgeführten Evaluierungen. In diesem FTB wird auch der Mid-term Report der österreichischen FTI-Strategie diskutiert. Schon das Wording zeigt, dass der Begriff der Evaluierung bei der Strategie selbst kein Thema ist.

### **Das FTI-Politik Umfeld – Aktuelle Entwicklungen**

Die kommende FTI-Strategie der Bundesregierung ist daher wichtig für die Ausrichtung der FTI-Politik-Evaluierung. Ein starker Impuls hinsichtlich Evaluierungsnotwendigkeiten (Planungshorizonte, ausreichende Budgetierung, Rahmen zur Bewertung missionsorientierter Programme etc.) wäre wünschenswert. Dies wurde auch in einigen Interviews im Kontext dieses Auftrages zum Ausdruck gebracht. Damit wäre auch ein Schritt verbunden, weg von v.a. legitimatorisch orientierten Programmevaluierungen hin zu einer umfassenden wirkungsorientierten Begleitung und ex post Bewertung, in der Evaluation in eine umfassendere FTI-Politik integriert ist.

Noch konkreter sind die Chancen und Möglichkeiten rund um die jüngst in Kraft gesetzte Forschungsfinanzierungsnovelle 2020<sup>5</sup> mit ihrem Kernstück, dem Forschungsfinanzierungsgesetz (FoFinaG). Dieses stellt einen bedeutenden Schritt in der Steuerung und Finanzierung der wichtigsten (außeruniversitären) Forschungseinrichtungen und Forschungsförderungseinrichtungen („zentrale Einrichtungen“) einschließlich der Förderagenturen AWS, FFG, FWF, CDG und OeAD dar. Mit jeder der Einrichtungen sind auf der Basis eines übergreifenden dreijährigen FTI-Paktes ebenfalls dreijährige Vereinbarungen zu schließen. Im Fall der insgesamt fünf Förderagenturen sind das Finanzierungsvereinbarungen, in denen die Budgets über diesen Zeitraum festgelegt und gesichert werden.

Die Ziele des FoFinaG (§ 1) setzen sich wie folgt zusammen: (1) *Langfristige, wachstumsorientierte Planungs- und Finanzierungssicherheit von FTI*, (2) *Strategische Ausrichtung und Steuerung von FTI*, (3) *Verwaltungsvereinfachung bei der Bereitstellung von Bundesmitteln ... sowie der Erhöhung der Effizienz in den Umsetzungsstrukturen sowie schließlich* (4) *Verbesserung von FTI-Leistungen und Analyse der erzielten Wirkungen*. In den Erläuternden Bemerkungen zur Novelle (EB) heißt es dazu: *Um in die Gruppe der europäischen und weltweiten Spitzenreiter, insbesondere auch in Bezug auf die Ergebnisse und Wirkungen von FTI (Outcome und Impact) vorzustoßen, bedarf es neben der Bereitstellung ausreichender Mittel auch einer Erhöhung der Effizienz und Treffsicherheit dieses Mitteleinsatzes. (S.1).*

Die Ziele des FoFinaG wie Planungssicherheit, strategische Ausrichtung oder Impact-Orientierung sind gut geeignet, im Rahmen der dreijährigen FTI-Pakte und Finanzierungsvereinbarungen operationalisiert zu werden. Die Pakte legen die *strategischen Schwerpunkte* über alle Einrichtungen – auch jene, die nicht zu den zentralen Einrichtungen gehören – hinweg fest und werden durch alle dafür zentral wichtigen Ressorts getragen (§ 2). Die Finanzierungsvereinbarungen mit den Förderungseinrichtungen legen dann die *konkreten Ziele und den Beitrag ... zur Umsetzung der Ziele des FTI-Pakts fest*; des Weiteren den Katalog der *Förderprogramme, ... Begleitmaßnahmen ... die Berichtspflichten der zentralen Forschungsförderungseinrichtung und Maßnahmen zur Qualitätssicherung und Kontrolle* (§ 7). Im § 7 sind dann auch die Mindestinhalte der Finanzierungsvereinbarungen festgelegt und im § 8 ein auf zentralen Indikatoren aufbauendes Monitoring der Performance der zentralen Einrichtungen, das dann Grundlage für Institutionen- und Systemevaluierungen bieten kann.

Für das kleine, unterfinanzierte, zeitlich kaum planen könnende Feld der FTI-Evaluierungen öffnet sich somit ein Möglichkeitsraum für größer dimensioniertes und strukturierteres Planen und Handeln. Offen ist dabei derzeit, wie dies geschehen soll und wird, weil das FoFinaG dafür nur den Rahmen aufspannt, selbst aber keine strukturellen oder organisatorischen Vorgaben macht. Das ist eine zentrale Frage, da sich in den nächsten Monaten eben entscheiden wird, welche Ebene auf der Basis gemeinsamer Planung und Vereinbarungen am Ende die Hoheit über die Programmgestaltung hat, wie diese Programme ausgestaltet werden und welche Regelungen, Mindestanforderungen, Verantwortlichkeiten und Budgets für Evaluierungen festgeschrieben werden.

Im FoFinaG gibt es keine eindeutigen Zuschreibungen, was Agenturportfolios, größere, missionsorientierte Programme und ähnliche Neuerungen in der Forschungsförderung betrifft; dies wird Sache der Strategie, der Pakte und der Finanzierungsvereinbarungen sein, die nun – teils parallel – strukturiert, vorbereitet und geschrieben werden. Das FoFinaG gibt aber mit seinen Zielen eine Richtung vor, die geeignet umzusetzen sein wird und in den EB heißt es, dass die Effizienz erhöht wird, weil die Finanzierungsvereinbarungen *eine große Anzahl an Einzelbeauftragungen ersetzen* (S. 2).

Vorgegeben ist auch das Ziel nicht nur in Europa vorne zu sein, sondern zu den weltweiten Spitzenreitern (EB, S.1) gehören zu wollen. Für entsprechende Verbesserungen gibt es in der Tat Chancen durch das Gesetz:

- Größere Interventionen zu thematischen oder strukturellen Aufgaben (siehe auch EB zu den Pakten, S. 4, *Themen sollen breit gefächert sein*), etwa missionsorientierte Interventionen, sind dadurch besser legitimier- und planbar. Die Chance lautet: Die Programme planen, die andere Länder schon haben, und die über eingegrenzte engere Themen und Interventionen hinausgehen, etwa größere *Missions*.

- Wo die Verantwortung für die Entwicklung von Programmen angesiedelt ist, lässt sich der gegenständlichen gesetzlichen Grundlage (dem FoFinaG selbst) nicht entnehmen, da diese Frage auch vom Förderungs- und Organisationsrecht beantwortet werden muss. Diesbezüglich gibt die Forschungsfinanzierungsnovelle eine gewisse Tendenz vor, indem für die AWS und die CDG die förderrechtlichen Voraussetzungen geschaffen wurden, Förderungen im eigenen Namen und auf eigene Rechnung durchführen zu können (§ 2 Abs. 2a AWSG und § 12a FTFG, FFG und FWF verfügen bereits über diese Möglichkeit). In den EB wird dies mit der wünschenswerten Erhöhung der Autonomie und Bürokratieabbau begründet (EB S. 14 und 20). Das heißt aber auch, dass die Chance nach wie vor besteht, im Sinn der FFG / aws Evaluierung (Bührer, et al., 2017) und ähnlicher Vorschläge den Agenturen zu überlassen, wie sie auf der Basis von Pakt und Finanzierungsvereinbarungen im einzelnen ihre Programme designen und welche Instrumente sie in welcher Form zum Einsatz bringen.
- Die Alternative dazu birgt freilich auch Chancen: Die Ministerien bestimmen weiterhin das Portfolio (stark mit), aber es gibt in Pakt und Finanzierungsvereinbarungen eine Festlegung auf eine geringere Zahl größerer und flexiblerer Programme, die auch geeigneter dafür sind, große gesellschaftliche Herausforderungen zu begegnen.
- Ebenso ist derzeit offen, wo die Verantwortung für Evaluierungen angesiedelt sein soll. Wenn die Ministerien weiterhin in Bezug auf die Programmgestaltung das entscheidende Wort sprechen, dann liegt auch die Evaluierungsverantwortung bei ihnen. Wenn die Programmportfolios überwiegend in den Agenturen entstehen und aufgesetzt werden, dann können die Evaluierungen entweder bei den Ressorts angesiedelt sein oder aber bei den Förderagenturen. Für letzteres findet sich ein Argument in den EB (S. 11): Die Mindestinhalte der Finanzierungsvereinbarungen enthalten Begleitmaßnahmen ... *Jedenfalls fallen darunter Dienstleistungen und Werkverträge zu Evaluierungen* ... (§ 7 Z. 3). Ein weiteres Argument, im Fall einer größeren Autonomie der Agenturen diesen auch die Vergabe von Programmevaluierungen zu übertragen, läge in einer Aufgabenteilung: Die Ressortebene übernimmt das in § 8 FoFinaG geregelte Monitoring und Evaluierung der Schlüsselindikatoren für den Gesamtfortschritt (Kennzahlen und Meilensteine, d.h. *Indikatoren, anhand derer die Erreichung der Ziele und Wirkungen der jeweiligen Finanzierungsvereinbarung gemessen werden kann*; § 7 Z. 12). Die Agenturebene hingegen übernimmt die Evaluierung der Einzelprogramme. Die Bestimmungen in § 7 Z. 12 bewirken jedenfalls, dass die Voraussetzungen für Evaluierungen geschaffen werden *müssen*, und dass infolgedessen entschieden werden *muss*, wer dafür die zentrale Verantwortung übernehmen soll.

Welchen Weg auch immer Ressorts und Agenturen beschreiten, die Pakte und vor allem die Finanzierungsvereinbarungen bieten eine große, vielleicht einmalige Chance zu zwei miteinander verbundenen Dingen: Erstens kann ein großer Schritt in der Definition, Entwicklung und Umsetzung von Programmen in Richtung Missionsorientierung, nachverfolgbaren Zielen, Größe, Flexibilisierung der Abwicklung und weiteren Eigenschaften gesetzt werden. Zweitens kann jeweils auf drei Jahre vorab festgelegt werden, welche Evaluierungsziele bestehen, und wer mit welchen (ausreichend ausgestatteten) Budgets wann welche Programme evaluiert.



Die FTEval wiederum sollte die Chance nutzen, für eine solche Strukturierung des Policy-Handelns und Aufwertung von Evaluierungen argumentativ und mit Vorschlägen aktiv zu werden.

### **Marktsituation und Akteure**

Wir können im Feld der österreichischen FTI-Evaluierungen von einem Markt sprechen, da es auf der Angebots- wie Nachfrageseite mehrere Akteure und regelmäßige Austauschbeziehungen gibt und weiters Preisbildung möglich ist.

Zugleich suggeriert das Bild des Marktes Größeres als eigentlich der Fall ist. Obzwar Zahl und Frequenz von FTI-Politikevaluierungen in Österreich seit einigen Jahren vergleichsweise hoch sind (Dinges & Schmidmayer, 2010; Streicher, Polt, & Unger, 2020), handelt es sich freilich um einen sehr kleinen Markt mit relativ wenigen Auftraggeber\*innen: drei Ministerien (BMBWF, BMK, BMDW) beauftragen meist direkt, und dazu kommen einzelne ihnen unterstehende Förderagenturen/Institutionen. Dazu kommen weitere Agenturen und regionale Akteure (Streicher, Polt, & Unger, 2020, S. 10). Die Zahl der Auftragnehmer\*innen ist hingegen größer, wächst und kann hinsichtlich Unternehmensgrößen, Rechtsform, Forschungsintensität und staatlichen Grundfinanzierungsanteilen als sehr heterogen bezeichnet werden. Zugleich ist es eine einstellige Zahl an institutionellen Akteuren, die an den meisten Evaluierungen als Auftragnehmer\*innen beteiligt sind. Hinzu kommt, dass für einige Organisationen FTI-Politik-Evaluierung ein wesentlicher Bestandteil des Umsatzes ist, für andere nur ein minimaler Anteil. Universitäre oder stark akademisch orientierte Einrichtungen sind kaum darunter, dem staatlichen Sektor zugehörige bzw. nahestehende Forschungsinstitute hingegen schon.

Die Marktstudie von (Streicher, Polt, & Unger, 2020) geht von einem jährlichen Auftragsvolumen von 750.000 Euro aus, was als Aussage schon für sich spricht. Überwiegend kommen bei Aufträgen aus Österreich österreichische Bieter zum Zug. Neben nationalen Aufträgen kommt für alle diese Organisationen insgesamt noch eine max. mittlere sechsstellige Summe pro Jahr aus internationalen Evaluierungen hinzu. Die meisten Evaluationsaufträge sind klein bis sehr klein, häufig unter 50.000 Euro. Oft sind sie auch nicht gänzlich ausfinanziert (Streicher, Polt, & Unger, 2020) was auch die hohe Zahl von über 100 Aufträgen in den letzten 10 Jahren relativiert. Das ist also der Markt (... wenigstens brauchen wir uns vor Google und Co nicht zu fürchten). Wenn auch neben Evaluierungen andere Instrumente wie Strategie- und Maßnahmenentwicklung, Expertenstudien oder Berichtsformate hinzutreten, ist klar, dass es für das knappe Dutzend einschlägiger Akteure schwer ist, Spezialisierungen zu entwickeln, zusätzliche Kompetenzen aufzubauen, in neue Felder vorzustoßen oder schlicht ein profitables Geschäft zu betreiben.

Der verstärkte Preiswettbewerb führt aus Sicht der Auftragnehmer\*innen zu einem Verzicht auf kostenintensivere Fragestellungen. Das deckt sich mit der Sicht der Auftraggeber\*innen: Es gibt hohen Wettbewerb bei kleineren Aufträgen, Zeitdruck, Budgetbeschränkungen in der Verwaltung (Streicher, Polt, & Unger, 2020), und kaum Koordination der öffentlichen Auftraggeber\*innen untereinander (Aussage Startworkshop). Das Timing der meisten Aufträge ist für die Auftragnehmer\*innen kaum planbar und oft nicht durchschaubar. Die meisten der Organisationen haben keine Basisfinanzierung. All diese Elemente tragen nicht dazu bei, dass österreichische Anbieter\*innen sich längerfristig in einer guten Position hinsichtlich großer internationaler Aufträge und damit verbundener Reputation befinden; dazu kommt, dass es einen verstärkten Wettbewerbsdruck auf internationaler Ebene gibt (Streicher, Polt, & Unger, 2020). Sie bieten auch keine günstige Grundlage für methodisch-

analytische Weiterentwicklungen im Feld der FTI-Politik-Evaluierung. Die Vergabe von einschlägigen Rahmenverträgen durch die EU-Kommission verschärft diese Problematik, da sie stärkere Akteure bevorzugt.

Wie sich der österreichische Markt auf der Auftraggeber\*innenseite entwickeln wird, ist schwer prognostizierbar. Die Zahl der Marktteilnehmer\*innen wird wohl in den nächsten Jahren eher noch steigen, auch werden wohl neue internationale Spieler wie etwa große Verlage hinzukommen, die zu Datenanalysten werden.

Abschließend ist noch zu fragen, ob es in der Vergangenheit Unterstützungsaktivitäten für die österreichische Evaluierungslandschaft gab. Die Antwort darauf könnte lauten: Immer wieder einmal, aber nicht in Form eines direkt darauf abzielenden Programms. Hilfreich war und ist die Unterstützung der Plattform FTEval durch stets mehr als ein Ressort, was immer wieder Schulungs-, Standardsetzungs- und Analyseinstrumente ebenso ermöglicht wie Sichtbarkeit durch ein internationales Konferenzformat. Hilfreich war das über viele Jahre hinweg finanzierte Programm TIP (technology – information – policy consulting) (Textbox 1) zur Analyse der FTI-Politik insgesamt, welches auch Evaluierungskompetenzen, Datenbestände und Sichtbarkeit stärkte.

Sehr wahrscheinlich hilfreich war auch das Programm der Dynamischen Qualitätssicherung (DQS) des Wissenschaftsressorts, mit dem kleineren außeruniversitären Forschungseinrichtungen in den Sozialwissenschaften Kompetenzaufbau, Weiterbildung und Luft zum Atmen ermöglicht wurden (**Textbox 2**, siehe auch Buchinger, Marchner, & Schoibl, 2007). Evaluierungskapazitäten standen dabei nicht im Programm, sehr wohl aber die Fähigkeiten weniger finanzstarker Einrichtungen, Kompetenzen und Methoden in der empirischen Forschung weiterzuentwickeln. Ebenfalls indirekt und hilfreich war die Systemevaluierung (Aiginger, Falk, & Reinstaller, 2009) als sehr gut dotiertes, auch größere Fragestellungen ermöglichendes Großereignis in der österreichischen Evaluierungslandschaft, mittlerweile über 10 Jahre her.

**Textbox 1:** TIP Programm (Peneder, 1999, S. 339)

**The Austrian “TIP” research programme**

“TIP is an innovation research and consulting programme based on an initiative of the Austrian Federal Ministry of Science and Transport and the Austrian Federal Ministry of Economic Affairs. It is carried out by the Austrian Institute of Economic Research (WIFO) in co-operation with the Austrian Research Centre Seibersdorf (ARCS). WIFO is in charge of directing and managing the programme.

The overall goal is to produce information and recommendations relevant to Austrian technology policy, based on analyses of technological change and its impact on the national system of innovation at the macro level (enterprises, public and private institutions and their interactions), the mesa level (structural analyses of the Austrian economy) and the micro level (analyses of firm behaviour).

Between 1993 and 1995, TIP evolved as a major research and consultancy programme designed to foster the knowledge base of Austrian technology policy through the provision of relevant analysis and the building of capabilities in the field. During this first term, strong emphasis was placed on cluster analysis of various kinds.

Based on the experience gained from the first phase of the programme, the core theme of the next phase (1996-99) was: "Ways into the Knowledge-based and Information Society Systems". The current work plan comprises the following modules:

- National System of Innovation (topics include a cluster case on multimedia and cultural content; inter-firm co-operation and innovation; the role of the service sector in the Austrian NIS; the role of the science system in technology development).
- Productivity, Growth and Employment (quantitative analysis of embodied interindustry flows; economic analysis of the relationship between R&D investment and productivity growth).
- Modern Regulation as an Instrument of Technology Policy (impact of the regulatory framework on innovative performance in the telecommunication and energy sectors).
- Consultancy for Technology Policy (providing stand-by capacity, e.g. for the monitoring of current developments in European Union industrial and technology policy)."

Das TIP Team entwickelte und betreute über viele Jahre auch den FTB und war wertvoll für Evaluierungsfragen, da diese in TIP-Arbeiten als Teil eines gesamthaften Policy-Ansatzes gesehen wurden.

## **Textbox 2: Programm Dynamische Qualitätssicherung GSK des BMWF<sup>6</sup>**

### **Programm Dynamische Qualitätssicherung GSK (2004-2009)<sup>7</sup>**

„Der Rat für Forschung und Technologieentwicklung (RFT) hat im Rahmen seiner Empfehlung für die Geistes- Sozial- und Kulturwissenschaften (GSK) vom 16.09.2003 die strukturellen Besonderheiten des außeruniversitären Sektors in diesen Wissenschaftsfeldern hervorgehoben und festgestellt, dass die dynamische Qualitätssicherung als Element der Forschungsoffensive besonderer Aufmerksamkeit bedarf.

Die Maßnahme „Dynamische Qualitätssicherung“ richtet sich an jene GSK-Forschungseinrichtungen, die vor allem anwendungsorientiert, fachübergreifend und wettbewerbsorientiert arbeiten, über Forschungsprojekte Drittmittel einwerben und sich gegenüber einer internationalen Konkurrenz bewähren müssen. Bei zahlreichen außeruniversitären GSK-Forschungsinstitutionen ist die Basisfinanzierung gemessen am internationalen Benchmarking unterkritisch. Eine temporäre Anschubfinanzierung durch zusätzliche finanzielle Mittel mit anschließender Evaluierung ist daher ein wichtiger Beitrag zur Qualitätssicherung der Arbeit dieser Institute und zur Erhaltung bzw. Erreichung von Spitzenleistungen.

#### **Zielsetzungen**

Mit dem Programm können Maßnahmen gefördert werden, die folgenden Zielsetzungen entsprechen:

- Erhöhung der Erfolgsquote bei Beteiligung an nationalen und internationalen Programmen;
- Aufbau und Verstärkung von nationalen und internationalen Vernetzungen und Kooperationen;
- Qualifizierung des wissenschaftlichen Personals und des wissenschaftlichen Nachwuchses; Förderung der wissenschaftlichen Karriere der Mitarbeiterinnen und Mitarbeiter;
- Erhöhung der Gleichstellung von Frauen und Männern in der Wissenschaft;
- Verbesserung der Forschungsinfrastruktur, insbesondere zur Bearbeitung von forschungsrelevanten Daten;
- Vertiefung der wissenschaftlichen Erkenntnisse aus Drittmittelprojekten;
- Erhöhung des Publikations-Impacts;
- Etablierung von Qualitätsmanagement, Organisationsentwicklung, Selbstevaluation.“

DQS lief von bis 2004-2009 und wurde im Zuge der Restrukturierungen des Förderwesens im Wissenschaftsministeriums ersatzlos gestrichen.

### **Akademisches Umfeld**

Die akademische Landschaft zur Erforschung und Analyse der FTI-Politik im Allgemeinen und von Evaluierungen im Speziellen ist im internationalen Vergleich in Österreich sehr schwach ausgeprägt. Es gibt praktisch keine universitäre Verankerung des Themenfeldes und nur wenige außeruniversitäre Gruppen verfügen über eine ausreichende Basisfinanzierung, um Methodenentwicklung und Wissensweitergabe systematisch durchführen zu können. In anderen europäischen Ländern ist die akademische Landschaft in dieser Hinsicht etwas aufgestellt. In Norwegen gibt es starke Akteure in der Innovationsforschung mit Evaluierungsbezug etwa in Trondheim bzw. Oslo, in Schweden akademische Gruppen in Stockholm, Lund und Göteborg, in Dänemark ebenfalls universitäre Gruppen in Kopenhagen und Aarhus, in Deutschland u.a. auf Universitäten in Berlin, Bonn und München oder in außeruniversitären Einrichtungen in Karlsruhe oder Hannover, in UK u.a. an den Universitäten von Manchester, Sheffield und Sussex, in den Niederlanden u.a. auf akademischer Ebene in Utrecht, Leiden, Maastricht und Twente oder auf Institutsebene etwa in Den Haag. Diese Liste lässt sich um französische, spanische, italienische, finnische, schweizerische Beispiele und viele weitere Orte und Organisationen ergänzen. Das ist eindeutig der internationale Standard.

Diese Einrichtungen forschen und publizieren, bilden Absolvent\*innen aus, betreiben PhD-Programme, entwickeln neue Methoden und gehen disziplinenübergreifende Kooperationen ein. Die wenigsten haben „Evaluierung“ als hauptsächliches Geschäftsfeld, aber das Thema spielt überall eine Rolle. Wieweit diese Wissenszentren auch nationale Politikgestaltung im Sinne ambitionierterer Modelle, Methoden und Reflexionsebenen beeinflussen, muss hier im Detail offenbleiben, Beispiele wie die Gruppe um J. Wilsdon in Sheffield oder das ISI in Karlsruhe lassen sich jedoch ohne Mühe finden. Der Diskurs und die Qualität des Angebots werden zweifellos befeuert. Große Initiativen wie das Transformative Innovation Policy Consortium (TIPC)<sup>8</sup>, das US Science of Science and Innovation Policy (SciSIP) Programm der NSF<sup>9</sup> oder das Research on Research Institute (RoRI)<sup>10</sup> haben schon kraft der dahinterstehenden staatlichen, privat-gemeinnützigen und akademischen Träger und Financiers eine Wirkung auf das Politik- und Fördersystem. Darauf wird in dieser Studie gleich noch beispielhaft einzugehen sein (siehe nächster Abschnitt).

In Österreich gibt es solche Einrichtungen nur in einem sehr geringen Ausmaß. Es gibt einzelne Professuren oder Forschungsgruppen mit gewissen thematischen oder geschäftlichen Nischen, aber nicht im vergleichbaren Ausmaß wie in den meisten anderen Ländern. Auffallend ist, dass die STS Community in Österreich vergleichsweise stark vertreten ist – mit Instituten an der Universität Wien<sup>11</sup> und an der Universität Klagenfurt<sup>12</sup>, sowie weiteren Personen am Institut für Technikfolgenabschätzung der ÖAW, am IHS sowie dem IFZ in Graz. Hochschulforschung gibt es mit Ausnahme einzelner Personen an der WU Wien, FH OÖ und der DUK so gut wie gar nicht, auch Ansätze aus ökonomischer Perspektive (BWL, VWL) sind auf wenige Personen (hier insbesondere WU Wien) beschränkt. Einige wenige Expert\*innen beschäftigen sich aus Sicht der Regionalforschung mit Innovation (Universität Wien, ÖAW). Eine rezente Studie des FWF et al. (Hesse, Matt, Reckling, Völker, & Possanner, 2019) kommt für den Bereich der Bildungsforschung, die man im weiteren

Sinne auch als Impulsgeber im Bereich FTI-Politik-Evaluierung sehen könnte, zu einem negativen Befund des Status quo.

Im kompetitiven akademischen Forschungsförderungsbereich zeigt sich, dass Disziplinen, die auch im FTI-Politik-Evaluierungskontext eine Rolle spielen könnten, nicht außerordentlich erfolgreich sind, wenn es um die Projektmittelinwerbung geht. Daten aus der FWF-Förderdatenbank zeigen, dass die empirischen Sozialwissenschaften in Österreich in der FWF-Einzelprojektförderung eher von geringerer Anzahl sind.

**Tabelle 1:** Auswertung der FWF-Förderdatenbank nach Anzahl der Einzelprojektförderungen in den genannten Disziplinen (nach dominanter Wissenschaftsdisziplin), 2010-2020.

Hauptsächliche Wissenschaftsdisziplin	Anzahl
Andere Sozialwissenschaften	6
Medien- & Kommunikationswissenschaften	5
Politikwissenschaften	11
Soziologie	44
Wirtschaftswissenschaften	57
<b>Gesamt</b>	<b>123</b>

Im Schnitt bedeutet dies 11-12 Projekte pro Jahr. Qualitative Methoden dominieren im sozialwissenschaftlichen Bereich mit Ausnahme der Wirtschaftswissenschaften. Aus diesen Projekten gibt es insgesamt vielleicht eine Hand voll Projekte, die mehr oder weniger in den Kontext der FTI-Politik-Evaluierung gebracht werden können. Analysiert man die Exzellenzpreise des FWF (START und Wittgenstein), zeigt sich, dass die Sozialwissenschaften noch weniger repräsentiert sind:

**Tabelle 2:** Auswertung der FWF-Förderdatenbank nach Anzahl der Wittgenstein und START-Preise, 1996-2019

	Wittgenstein	START
Gesamtzahl der vergebenen Preise	37	146
mit zumindest tw. sozial- und wirtschaftswissenschaftlichen Anteil	5	7
im Kern Sozial- und Wirtschaftswissenschaft	3	2

Die akademische Reputation bei Evaluierungen spielt aus Sicht der Auftraggeber\*innen auch kaum eine Rolle [Interview Auftraggeber\*innen] und es werden kaum akademische Einrichtungen beauftragt: Nicht die heimischen, weil es die eben kaum gibt und nur selten die internationalen, weil sie nur für einzelne Fragen wie bibliometrische Evaluierungen oder Karriereverläufe nachgefragt werden. Ein Bereich derzeit noch schwacher, aber wahrscheinlich zunehmender Bedeutung sind die Computerwissenschaften / Data Sciences, die über Big Data / AI / Machine Learning / Complexity Science neue Tools auch für die Bewertung von staatlichen Initiativen und ihrer Wirksamkeit haben. Hier sind in Österreich akademische Kompetenzen vorhanden, jedoch gibt es weder vom Evaluierungsgegenstand noch vom Angebot noch von der Nachfrage her derzeit starke Anknüpfungspunkte und Signale.

*Kurze Einschätzung:* Im Hochschulumfeld gibt es kaum gelebte Verbindungen zu tatsächlichen Evaluierungen in der Praxis weder aus den Sozialwissenschaften noch in den Computerwissenschaften. Das Themenfeld Evaluation sowie Forschungs- und Innovationspolitik steht nicht im Fokus der Themensetzungen von Hochschulen. Auch die geringen Marktgröße bei Evaluierungen spielt hier eine Rolle. Die geringe Marktgröße und fehlende Freiheiten in den kleinen Auftragsprojekten, die man nutzen könnte, um daraus einen akademischen Mehrwert in Form von Publikationen zu generieren, machen das Feld aus einer akademischen Sicht nicht sehr attraktiv. Umgekehrt spielt die grundlagenorientierte Drittmittelförderung (z.B. über den FWF, siehe auch Tabelle 1 und 2) im gesamten Bereich der Policy-Forschung und generell der empirischen Sozialwissenschaften keine große Rolle und damit fehlt eine gemeinsame Förderquelle, um etwa Kooperationen von universitärer und außeruniversitärer Forschung angehen zu können und einen Wissenstransfer im Methodenbereich zu ermöglichen. Hinzu kommt natürlich, dass die Orientierungspunkte für den akademischen Bereich (hochrangige internationale Publikationen) und dem FTI-Politik-Bereich (Erstellen von Berichten nach den Vorgaben der Auftraggeber) eine Zusammenarbeit erschweren können. Das alles klingt nicht nach *capacity building*.

### **Initiativen aus anderen Ländern, die uns etwas sagen können**

#### **Förderprogramm des BMBF (Deutschland)**

Programm zur Förderung der *Weiterentwicklung der Indikatorik für Forschung und Innovation*<sup>13</sup> mit bislang zwei Ausschreibungen 2017 und 2018/19. Hier handelt es sich um ein Förderprogramm im erweiterten Rahmen der deutschen HighTech-Strategie mit ihrem breiten Innovationsbegriff. Es dient zur Unterstützung der FTI-Politik, indem eben konkrete Indikatoren entwickelt werden [Interview ISI]. Das betrifft einerseits Indikatoren in bereits bekanntem Territorium, andererseits solche, die mit neueren / noch nicht gut abgesicherten Entwicklungen wie Dienstleistungsinnovationen oder „Open-“, „Sustainable-“ oder „Responsible-“ Themen zu tun haben. Es sollen mit dem Programm Forschungslücken geschlossen werden und dabei auf internationale Entwicklungen wie die Überarbeitung der OECD-Manuale Bedacht genommen werden. Die Empfänger\*innen sind oft interdisziplinäre und interinstitutionelle Teams aus universitären und außeruniversitären Forschungseinrichtungen, darunter mehrere Verbundprojekte.

Der Fokus liegt auf Indikatorenentwicklung, aber die Grenzen zu einer allgemeineren Methodenentwicklung scheinen freilich fließend zu sein, angesichts von Fragestellungen in der ersten Ausschreibung, wie: „*Welche methodischen Neuerungen und welche neuen empirischen Zugänge können zur Verbesserung von Wirksamkeitsanalysen und zum Innovations-Monitoring beitragen?*“ Oder: „*Wie können Karrierewege in der Forschung und deren Bedeutung für Innovationen erfasst werden?*“ (Ausschreibungstext erster Call<sup>14</sup>). Eines von mehreren angesprochenen Themenfeldern der zweiten Ausschreibung betrifft auch die Entwicklung von Input-, Output- und Outcome- „*Indikatoren zur Unterstützung von Evaluationen von FuE-Förderungen.*“ (Ausschreibungstext zweiter Call<sup>15</sup>). Damit sind explizit etwa auch Kontrollgruppenansätze angesprochen. Bei all der Nähe zu Methodenfragen insgesamt ist jedoch auch zu betonen, dass die Initiative vom Statistikreferat des BMBF kommt, um irgendwie die neue Realität empirisch und statistisch dingfest machen zu können [Interview vdi-vde].

Die Förderungen sind relativ gut dotiert und können im Fall der Verbundprojekte in einem Dreijahresrahmen bis 450.000 € gehen. Die Ausschreibungen werden vom vdi-vde als Projektträger durchgeführt, die Förderempfehlungen von einer deutschsprachigen, aber in diesem Rahmen auch internationalen Jury vorgenommen. In dieser Jury sind auch mehrere österreichische Expert\*innen vertreten gewesen. Die Bandbreite der in den 17 bislang geförderten Projekten ist sehr groß und reicht von IT-gestützten Verfahren über Mikrodaten bis hin zu komplexen Netzwerken und ihren Interaktionsmustern.

**Tabelle 3:** 1. Ausschreibung 2017 des BMBF Förderprogramms *Weiterentwicklung der Indikatorik für Forschung und Innovation*<sup>16</sup>

Titel des Projekts	Projektleiter*in	Institution	Projektwebsite
Textdaten-basierte Output-Indikatoren als Basis einer neuen Innovationsmetrik	Georg Licht	Zentrum für Europäische Wirtschaftsforschung (ZEW), Justus-Liebig-Universität Gießen	<a href="https://www.zew.de/forschung/textdaten-basierte-output-indikatoren-als-basis-einer-neuen-innovationsmetrik">https://www.zew.de/forschung/textdaten-basierte-output-indikatoren-als-basis-einer-neuen-innovationsmetrik</a>
Innovationsindikatorik für den Doing-Using-Interacting-Mode von KMU	Kilian Bizer	Georg-August-Universität Göttingen	<a href="https://indui.wordpress.com/">https://indui.wordpress.com/</a>
InnoMetrics	Matthias Rüdiger	RWTH Aachen	<a href="https://www.time.rwth-aachen.de/cms/TIME/Forschung/Forschungsprojekte/~davcj/Innometrics/">https://www.time.rwth-aachen.de/cms/TIME/Forschung/Forschungsprojekte/~davcj/Innometrics/</a>
Initiative für Offene Wissenschaft und Innovation	Marte Kessler	Gemeinnützige Gesellschaft für Wirtschaftsstatistik GmbH im Stifterverband	<a href="https://www.stifterverband.org/initiative-offene-wissenschaft-innovation">https://www.stifterverband.org/initiative-offene-wissenschaft-innovation</a>
Technologische Räume - Evolution, Potenziale und politische Implikationen - TechSpace	Uwe Cantner	Universität Bremen	<a href="https://www.microtheory.unijena.de/research/techspace">https://www.microtheory.unijena.de/research/techspace</a>
Investitionen in neue Formen von wissensbasiertem Kapital (INFOWIK)	Thomas Niebel	Westfälische Hochschule Gelsenkirchen Bocholt Recklinghausen	<a href="https://www.zew.de/forschung/projekte/investitionen-in-neue-formen-von-wissensbasiertem-kapital-in-fowik">https://www.zew.de/forschung/projekte/investitionen-in-neue-formen-von-wissensbasiertem-kapital-in-fowik</a>
Indikatorik Soziale Innovation / IndiSI	Judith Terstriep	Zentrum für Europäische Wirtschaftsforschung (ZEW)	<a href="https://www.iat.eu/forschung-und-beratung/projekte/2018/indisi-indikatorik-soziale-innovation.html">https://www.iat.eu/forschung-und-beratung/projekte/2018/indisi-indikatorik-soziale-innovation.html</a>

**Tabelle 4:** 2. Ausschreibung 2018/19 des BMBF Förderprogramms *Weiterentwicklung der Indikatorik für Forschung und Innovation*<sup>17</sup>

Titel des Projekts	Projektleiter*in	Institution	Projektwebsite
Netzwerke innovativer Unternehmen (NETINU)	Christian Rammer	Technische Universität München, Leibniz-Zentrum für Europäische	<a href="https://www.zew.de/forschung/netzwerke-innovativer-unternehmen-netinu/">https://www.zew.de/forschung/netzwerke-innovativer-unternehmen-netinu/</a>

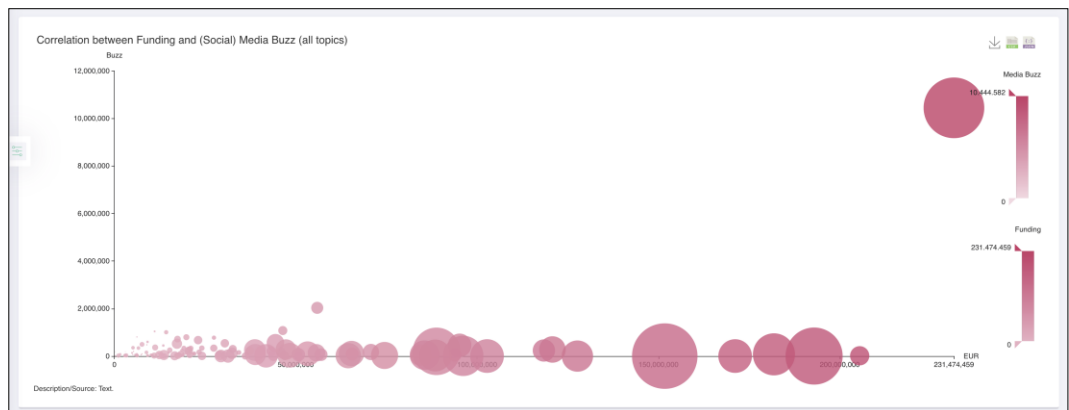
		Wirtschaftsforschung Mannheim	
Open-Access-Informationen für alle	Anja Oberländer	Technische Informationsbibliothek Hannover	<a href="https://open-access.network/startseite">https://open-access.network/startseite</a>
Ganzheitliche Erfassung von Innovations-Ökosystemen: Erfassung formeller, informeller und nicht-kodifizierter Austauschbeziehungen in Innovationssystemen	habil. Henning Kroll	Fraunhofer Institut für System- und Innovationsforschung Karlsruhe	<a href="https://www.isi.fraunhofer.de/de/competence-center/innovations-wissensoekonomie/projekte/innovations-oekosysteme.html#tabpanel-18415641231">https://www.isi.fraunhofer.de/de/competence-center/innovations-wissensoekonomie/projekte/innovations-oekosysteme.html#tabpanel-18415641231</a>
UseAltMe	Simone Fühles- Ubach,  Dirk Tunger	Technische Hochschule Köln	<a href="https://www.th-koeln.de/en/information-science-and-communication-studies/usealtme_68578.php">https://www.th-koeln.de/en/information-science-and-communication-studies/usealtme_68578.php</a>
PeerInnovation – Nachhaltige Haushaltsinnovation in kollaborativen Netzwerken	Jan Peuckert	Institut für ökologische Wirtschaftsforschung, Technische Universität Berlin	<a href="https://www.ioew.de/projekt/peerinnovation_nachhaltige_haushaltsinnovation_in_kollaborativen_netzwerken">https://www.ioew.de/projekt/peerinnovation_nachhaltige_haushaltsinnovation_in_kollaborativen_netzwerken</a>
Indikatoren zu regionalen Wissenstransferstrukturen für Nachhaltige Entwicklung	Daniel Hirschmann, Daniel Feser, Kilian Bizer	Georg-August-Universität Göttingen, Hochschule Darmstadt	<a href="https://www.uni-goettingen.de/de/616304.html">https://www.uni-goettingen.de/de/616304.html</a>
WISIH – Soziale Innovationen aus Hochschulen	Isabel Roessler	CHE Gemeinnütziges Centrum für Hochschul- entwicklung	<a href="https://www.che.de/projekt/wisih-soziale-innovationen-aus-hochschulen/">https://www.che.de/projekt/wisih-soziale-innovationen-aus-hochschulen/</a>

### Beispiel Projekt mit Beteiligung des Fraunhofer ISI (Deutschland)

Unter Beteiligung des ISI Fraunhofer (verantwortlich für den konzeptionellen Rahmen) wurde ein EU-Projekt durchgeführt, das neue Methoden zur Bewertung der Performanz und Auswirkungen von Forschung und Innovation entwickelt hat. In *Data4Impact* (Big Data approaches for improved assessment of the societal impact in the Health, Demographic Change and Wellbeing Societal Challenge)<sup>18</sup> ging es darum, auf Big Data Basis (mehr als 13.000 EU Projekte aus dem Gesundheitsbereich wurden analysiert) ein Set von Indikatoren zu schaffen, das neben Input und Output auch den akademischen, ökonomischen und gesellschaftlichen Impact umfasst. Für den Impact wurden insgesamt 10 Indikatoren geschaffen wie beispielsweise „Impact on public health“, der die Zitationen eines Projekts in klinischen Guidelines zählt.



**Abbildung 1:** Darstellung der gesellschaftlichen Impacts im Projekt *Data4Impact* (Ausschnitt aus der Website)



*Kurze Einschätzung:* Das BMBF fördert mit einigen Millionen Euro ein ordentlich dotiertes Programm. Auch wenn die BMBF-Initiative ähnlich wirkt wie die Aufgabenstellung für unsere Studie, gibt es doch zwei große (miteinander verkoppelte) Unterschiede: Erstens ist das Ziel des BMBF, konkret Indikatoren für die statistische Erfassung und für die Steuerung der FTI-Politik zu gewinnen: „*Forschung und Innovation transparenter machen*“ ist die Devise [Interview vdi-vde]. Zweitens ist das Programm nicht dazu da, Institutionen zu unterstützen, um marktfähig zu bleiben. Das hat auch seine Gründe: Organisationen wie ISI oder IFQ sind ordentlich dotiert und die universitären Einrichtungen haben auch alle ihren Grundfinanzierungsanteil. Ein Nebeneffekt des Programms wird auch sein, dass andere Länder von den Ergebnissen profitieren können und werden (siehe Kapitel 4 Empfehlungen). Für unsere konkrete Aufgabe heißt das: Keine Verdopplung des Programms, aber Lernen aus demselben und in Österreich starke Konzentration auf die Datenfrage als Grundlage für Indikatoren. Indikatoren sind (sofern es sich nicht um ad hoc Indikatorik handelt, um in einer Forschungsprojekt eine Erkenntnisfrage zu beantworten) eine Frage grundlegender Standards / Infrastruktur, die v.a. politisch beantwortet werden muss.

### **Transformative Innovation Policy Consortium (UK / Niederlande / weltweit)**

Das Transformative Innovation Policy Consortium (TIPC) ist am SPRU der Universität Sussex angesiedelt, mit einem starken Link in die Niederlande verbunden. Finanziert wird es in einem Fünfjahresprogramm durch eine globale Allianz von nordeuropäischen (RCN, Vinnova, Business Finland / ex TEKES), afrikanischen und lateinamerikanischen Innovationsförderagenturen. Weitere Partner sollen bzw. werden folgen. Hinter dieser Initiative steht die Idee, dass FTI-Politik historisch seit 1945 jeweils einem dominanten „Frame“ folgt(e), von dem sich auch Politikverständnis, Instrumente und Interventionen sowie Bewertungsmaßstäbe und Evaluierungsformen ableit(et)en. Die älteren Frames bleiben auch neben dem neuen weiter bestehen.

Über die beiden großen Rahmenerklärungen des linearen Modells der Nachkriegszeit und der Innovationssysteme seit den 1980er Jahren hinaus versucht TIPC (wie viele andere Akteure auch), für den aktuell entstehenden dritten Frame ein Modell und Praktiken zu entwickeln. Dieser Rahmen ist durch dichte Interaktionsmuster, Missionsorientierung, die SDGs und Nachhaltigkeitsziele, Verkopplung mit Sektorpolitiken, Einbindung von Praxispartnern und insgesamt den Versuch von systemischen Änderungen geprägt. Stichworte

dafür sind *directionality, societal goals, system level impacts, learning and reflexivity, conflict / consensus, inclusiveness* (Schot & Steinmueller, 2018).

Ziel ist die Entwicklung und Erprobung von Konzepten und Instrumenten für *transformative policies* durch empirische Forschung, Evaluierung, Training, Aufbau von Kapazitäten und Public Outreach: „... *encompasses experimentation with robust transformative evaluation techniques that support a re-make of systems ... to meet social, environmental and economic challenges ...*“<sup>19</sup> Der experimentelle Charakter steht im Vordergrund. Niemand weiß wie das mit der Systemtransformation geht. Alle wissen, dass Komplexität, Akteursvielfalt und Interessensgegensätze ihresgleichen suchen, wenn es um den Umbau zu einer nachhaltigen Gesellschaft geht und dass es an fast jeder Stelle dieser Transformation zu schwerwiegenden Fehlsteuerungen, Rebound-, Gegen- und ungeplanten Effekten kommen kann und wird. Alle (gut, fast alle) wissen, dass diese Übung gelingen muss und daher durch Suchen und Versuchen die Erfahrung und Sicherheit gewonnen werden muss. Im Evaluierungsbereich betrifft das neben Konzeptentwicklungen vor allem formative Evaluierungen. Konkret geschieht das etwa durch Verkopplung mit konkreten Design- und Evaluierungsprozessen, siehe dazu unten das schwedische Beispiel (unten, Seite 25).

*Kurze Einschätzung (mit einer kleinen Abweichung vom Thema):* Die Initiative transportiert im Gegensatz zu früheren Innovationsparadigmen eine sehr grundsätzliche Unsicherheit, was Hebelpunkte, Akteurszusammenspiel, Kausalitäten und dergleichen betrifft. In früheren Frames hat es geheißt: *Wir kennen viele Zusammenhänge auch nicht oder können sie kausal nicht belegen, aber macht Euch keine Sorgen. Wir haben dennoch ein starkes Erklärungsmodell und wenn ihr uns nur lässt und finanziert, wird die Zukunft gut und sie wird gelingen.* Das hat zu gewaltigen Fortschritten geführt und zu einer globalen Durchdringung mit dominanten Technologiemustern. Ihre Wirkungen – von Gesundheitsrisiken über die Klimakrise bis zum Überwachungskapitalismus – lassen sich aber nicht mit den gleichen Mitteln beantworten, mit denen die Ursachen erzeugt worden sind. Daher ist der suchende Ansatz mehr als gerechtfertigt, auch wenn es tausend Gefahren gibt, sich im Wald zu verlaufen und in den Spalt zwischen großen Fragen und überschaubaren, heterogenen und kleinen Experimenten hineinzufallen. Für die Evaluierungscommunity ist der Hauptpunkt wohl der, dass der Weg über Sektorpolitiken und deren Daten sowie über das Verständnis neuer, größerer Akteurskonstellationen interessante methodische und empirische Neuerungen erlauben kann. In Österreich kennen wir diesen Weg mangels Versuche und Programmen noch kaum, es sind noch keine ernsthaften Mission-Programme aufgesetzt, die ebenso eine Fehlstelle darstellen wie übergreifende Evaluierungen zu Grand Challenges.

### **Schweden: Eine sehr große Evaluierung**

Schweden ist ein Land mit einer gut entwickelten FTI-Politik und Evaluierungspraxis. Dabei stechen etwa Versuche namentlich der Agentur VINNOVA und ihrer Vorläuferinnen hervor, Wirkungen von Initiativen und Programmen über lange Zeiträume hinweg zu identifizieren und zuzuordnen sowie Wirkungsstudien auch miteinander zu vergleichen (siehe etwa (Elg & Hakansson, 2012)). In einzelnen Themen wie Mobilfunk oder Transport konnten so schon Erfahrungen gesammelt werden. Nunmehr steht das Land und sein Innovationsförderer vor der Herausforderung, zahlreiche breite, missionsorientierte „change“-Programme nicht nur aufzusetzen,<sup>20</sup> sondern den Wandel, die Erfüllung der Mission und die Wirkung auf große Akteursgruppen auch mit Evidenz aus Evaluierungen zu unterlegen.

Unter den wichtigsten einschlägigen Programmen in Schweden zielt das 2012 auf eine lange Zeitperiode eingerichtete, große Strategic Innovation Programme (SIP oder auch SIO abgekürzt) auf größerflächige Veränderungsprozesse in thematisch wichtigen Feldern (Grand Challenges) durch intensive Kooperation und Interaktion zwischen Wissenschaft, Wirtschaft und Gesellschaft (OECD, 2016, S. 107ff.). Dies ging mit der Entwicklung neuer Instrumente und neuer Erfolgskriterien, von *Strategic Research Agendas* sowie mit hoher Delegation von Management einher. Zahlreiche große SIPs wurden ausgewählt und eingerichtet. Die Größe und Neuartigkeit des Programms stellt für die Förderagentur eine beträchtliche Herausforderung dar.

**Textbox 3:** Innerhalb der SIP sind die Projekte eigene Sub-Programme: Beispiel *Viable Cities*

#### **Viable Cities: Struktur und Schwerpunkte, Verhältnis zum Gesamtprogramm**

Viable Cities ist ein gefördertes Vorhaben im SIP und für sich genommen wiederum ein strategisches Innovationsprogramm, mit Schwerpunkt auf intelligenten, nachhaltigen Städten. Um ein Gefühl für die Größe zu bekommen: Allein Viable Cities als ein „Subprogramm“ hat für 12 Jahre ein Budget von 100 Mio. € und vergibt diese Mittel teils in vom Konsortium selbst organisierten Wettbewerben. Die Mission von Viable Cities ist es, den Übergang zu klimaneutralen Städten bis 2030 zu beschleunigen und ein gutes Leben für alle innerhalb der Grenzen des Planeten zu führen. In diesem Sinn ist Viable Cities ein Katalysator für **neue Formen der Zusammenarbeit zwischen Städten, Industrie, Wissenschaft, Forschungsinstituten und der Zivilgesellschaft**. Dies soll mobilisieren, um die Arbeitsweise unserer Städte im Einklang mit den nationalen und internationalen Umwelt- und Klimazielen zu ändern. Die Laufzeit ist 2017-2030 und wird alle drei Jahre evaluiert. Es wird mit Unterstützung von Vinnova, der schwedischen Energieagentur und Formas umgesetzt.

Die vier Schwerpunkte des Programms werden unter klarer Einbeziehung der Bürger umgesetzt: Lebensstil und Konsum, kombiniert mit Planung und bebauter Umwelt, Mobilität und Zugänglichkeit sowie integrierte Infrastruktur sind die größten Energieverbrauchssektoren in einer Stadt. Fünf konkrete Themen konzentrieren sich auf die Aktivitäten und stellen sicher, dass Technologieentwicklung und Wissensentwicklung mit starkem Einfluss der Bürger und mit Menschen im Zentrum stattfinden: Test beds & Living Labs, Innovation und Unternehmertum, Finanzierung und Geschäftsmodelle, Kontrolle, Intelligenz, Cybersicherheit und Ethik. Aktivitäten und Initiativen von Viable Cities finden an den Schnittstellen zwischen Schwerpunktbereichen und Themen statt.<sup>21</sup> Dazu gibt es variierende Ausschreibungen im Rahmen von Viable Cities für kollaborative konkrete Projekte, die selbst auch einen missionsorientierten Charakter haben sollen. Im SIP Programm gibt es somit drei Ebenen, die den Transformationsanspruch stufenweise konkretisieren: Das Gesamtprogramm, die einzelnen Förderinitiativen wie eben Viable Cities und innerhalb derselben konkrete Ausschreibungen für Einzelvorhaben.

#### **Ziel dieser missionsorientierten Förderinitiative**

Vision: Schweden inspiriert und spielt eine führende Rolle bei der Energiewende und dem Klimawandel durch intelligente und nachhaltige Städte. Mission: Wir werden den Übergang zu klimaneutralen Städten bis 2030 beschleunigen und ein gutes Leben für alle innerhalb der Grenzen des Planeten führen.

Ziel ist es, die Fähigkeit schwedischer Städte zu stärken, eine treibende Kraft beim Übergang zu klimaneutralen Städten zu sein. Dies geschieht auf der Grundlage eines ganzheitlichen Ansatzes für eine nachhaltige Stadtentwicklung mit Menschen im Mittelpunkt, bei denen Digitalisierung und bürgerschaftliches Engagement wichtige Voraussetzungen sind. Dabei steht ein missionsorientierter Ansatz im Vordergrund<sup>22</sup>. Eingebettet ist diese Arbeit in internationalen Kooperationen; Viable Cities will ein Hub in einem globalen Netzwerk intelligenter, nachhaltiger Städte sein.

### Zwischenevaluierung<sup>23</sup>

Nach drei Jahren erfolgte eine erste von vier (Zwischen-)Evaluierungen von Viable Cities bei einer Laufzeit von 12 Jahren. Mit dieser Bewertung im Frühjahr 2020 sollte unter anderem festgestellt werden, wie gut das Programm eingerichtet wurde, und die Stärken und das Verbesserungspotenzial des Programms hervorgehoben werden. Positiv hervorgehoben wird in der Evaluierung vor allem der ausgeprägte ganzheitliche Ansatz („Viable Cities hat eine moderne Sicht auf Innovation und die Rolle, die der Staat spielen kann, übernommen“) sowie der Bezug zur EU und Policy-Ebene. Verbesserungspotenzial wird gesehen für die Klärung der Positionierung und mehr Zusammenarbeit der Viable Cities Initiative in Bezug auf andere Initiativen (national sowie international). Es wird auch vorgeschlagen sich breiter an die umsetzenden Akteure zu wenden und auch Akteure identifizieren und in größerem Umfang einbeziehen, die das derzeitige Regime in Frage stellen („Die Methoden zur Erfassung von Stakeholdern und unterstützungsbedürftigen Eigentümern sind trotz des innovativen *transition management* weitgehend konventionell“). Die Rolle und Beteiligung von Unternehmen muss geklärt und gestärkt werden. Das *transition management* wird als Grundlage für das Programm beschrieben, aber seine Aktivitäten sollten klarer gesehen werden. Parallel zu guten Initiativen wie dem *transition lab* sollten andere Aktivitäten zur Umsetzung im Einklang mit dem *transition management* eingeleitet werden. Der Gefahr der „Ausfransung“ des Programms muss entgegengewirkt werden.

Die Antwort darauf ist auch eine sehr große Evaluierung. Über mehrere Jahre hinweg sind 3 Mio. € (nicht Kronen, wir haben nachgefragt, [Interview Vinnova]) für die Evaluierungen der einzelnen SIPs und des Programms insgesamt reserviert. Diese „*biggest evaluation ever in Sweden*“ [Interview Vinnova] soll verstehen lernen helfen, wie Missionen aufzusetzen sind, wie sie wirken und wie gelernt werden kann, Schritte der Systemtransformation zu bewältigen. Nicht nur die Förderungen, auch andere Politikinstrumente wie Vergabepolitik stehen dabei im Fokus. Dazu gibt es auch eine Zusammenarbeit mit SPRU / TIPC (siehe oben) oder mit M. Mazzucato von der UCL, auch verbunden mit dem internationalen Beirat der Agentur. Diese Evaluierung wird als ein Labor für Methodenentwicklung betrachtet, namentlich für begleitende und formative Schritte, denn „*system transformation is no case for ex post evaluations*“ [Interview Vinnova]. Die Tatsache, dass schwedische Agenturen selbst die Evaluierungen beauftragen (auch weil sie in den kleinen Ministerien gar nicht verarbeitet werden können), wird als sehr wichtig angesehen. Eine Nebenbemerkung noch: Vinnova fördert ganz allgemein auch FTI-Politikanalysekapazitäten an den Universitäten Lund und Göteborg mit mehrjährigen Rahmenverträgen.

*Kurze Einschätzung:* Die schwedische Innovationspolitik traut sich selbst etwas zu, das war in den letzten 20 Jahren nicht immer so. Sie machen große Programme und haben Veränderungsziele. In diesem Frame – und im Gegensatz zu Österreich – wird nicht immer a priori davon ausgegangen, dass die Unternehmen vorab und alleine am besten wissen was zu tun ist und daher gleich die Subvention erhalten sollen. In diesem Frame ist auch klar, dass mit der staatlich gestützten Preisgestaltung für die Ingenieursstunde allein werden Herausforderungen wie die Klimakrise nicht bewältigt werden. Solche großen, komplexen Programme brauchen viele Ressourcen, um zu steuern, zu lernen und zu verstehen. Sie sind daher – gegeben, dass es große, lernoffene Evaluierungen gibt – eine gute Brutstätte für Methodenentwicklung. Große Evaluierungen als *innovative procurement* des Staates. Für unsere konkrete Aufgabe heißt das: Ambitioniertes Staatshandeln, entsprechende Programme, Lernoffenheit und größer dimensionierte Evaluierungen erlauben die Entwicklung und das Testen neuer Evaluierungsmethoden.

## **International: Research on Research Institute (RoRI)**

RoRI<sup>24</sup> ist eine sehr junge Initiative, die von großen philanthropischen und staatlichen Förderinstitutionen ausgeht. Der Wellcome Trust, die Sloan Foundation, Howard Hughes Medical Institutes, die Volkswagenstiftung oder die Chan Zuckerberg Initiative stehen etwa für die erste Kategorie, NWO, SNSF und auch der FWF für die zweite. Weitere globale Partner sollen dazu kommen. Die Universitäten Sheffield (Gruppe um James Wilsdon) und Leiden (CWTS) sind das akademische Rückgrat. Der Name ist Programm: Die Beteiligten wollen mehr über ihr Handeln, dessen Wirkungen und die relevanten Bezugssysteme wissen, um Steuerbarkeit, *accountability* und Impact-Orientierung zu erhöhen. Im Gegensatz zu TIPC handelt es sich hier eher um Akteure aus der Wissenschaftsförderlandschaft und um deren ungelöste Fragen, die von konkreten Evaluierungsfragen abstrahieren und grundsätzliche Zusammenhänge in der Wissenschaft klären wollen: Auswahlprobleme (Randomisierung), direkte und indirekte Wirksamkeit der Förderungen und Programme, Transparenz, Wiederholbarkeit und ethische Standards in der Forschung etc. Ein konkretes Projekt beschäftigt sich mit der (ungeklärten) Frage, ob Humanressourcen-Förderung den größten Social Impact erzeugt [Interview CWTS].

RoRI ist auch nicht die erste Initiative, die versucht diese Elemente im Wissenschaftssystem zu stärken; vor vielen Jahren schon hat es in den USA die Science of Science Initiative der Regierung gegeben. Bei RoRI dürfte eine starke Energie von den häufig sehr großen und vielfach stark gewachsenen philanthropischen Einrichtungen ausgehen, von denen sich einige mit bisherigen Erklärungs- und Wirkungsmustern nicht gleich zufriedengeben. Ähnlich verhalten sich avanciertere staatliche Förderer. Alle haben das gleiche Problem: Der Druck ist überall höher, das Wissenschaftssystem größer, komplexer geworden und steht mehr in der Auslage als je zuvor.

RoRI soll eine Reihe von Analyse- und Forschungsströmen bündeln und erst für die Mitglieder und dann für die weitere Welt verfügbar machen. Wichtig dabei ist die Nachfrageorientierung, sprich die Förderinstitutionen bestimmen stark die Forschungsagenda mit und sind in der Entwicklung und Testung aktiv. Das Institut arbeitet verteilt und es gibt mehrere Arbeitsstränge, an denen in Partner Labs jeweils einige Mitglieder teilnehmen. Neben Kernprojekten will RoRI sich auch in andere Konsortien einklinken oder als Berater auftreten. Dazu gibt es – auch verteilt – Arbeiten an Datensammlungen, Big Data Analysis etc., etwa beim CWTS oder bei Förderern wie Wellcome.

Die gemeinsame Entwicklung und Testung neuer Evaluierungsinstrumente und Indikatoren ist eines der Themen und nimmt eine wichtige Rolle ein. Bei den 15 groben Themen kommt Evaluierung aber auch an vielen anderen Stellen vor, von Bibliometrie über Gendergerechtigkeit, Karrieren bis hin zu Kooperationen. In den drei großen Themen *decisions*, *careers*, *cultures* geht es vor allem um neue Indikatoren, die Bestätigung möglichst objektiven Handelns und um die methodisch bessere Erfassung von Impacts einschließlich der weiteren Entwicklung von Geförderten.

*Kurze Einschätzung:* Große Fragen bedürfen großer Anstrengungen. Die beiden Themen „Impact“ und „Accountability“ rücken international stärker in den Vordergrund und es werden entsprechende Ressourcen von Stiftungen und Förderern gewidmet. Hier die Augen zu verschließen und nur kleine, gelegentliche und weniger strenge Fragen zu stellen, wird für ein staatliches System, das viel auf Förderung setzt, zukünftig schwieriger werden. Die Teilnahme etwa des FWF an Initiativen wie RoRI ist daher als sehr positiv einzuschätzen.

RoRI ist ein Indikator für eine Reihe von Veränderungen in der FTI-Politik: Zum einen gibt es hier Kooperationen zwischen privaten und öffentlichen Akteuren bzw. einer Vielfalt von unterschiedlichen Organisationstypen, die hier zusammenarbeiten. Zum zweiten ist das *center of gravity* Daten, die die unterschiedlichen Partner einbringen. Epistemischer und ökonomischer Mehrwert wird durch den Zugang und die Aufbereitung der Daten generiert (und nicht durch eine spezifische Methodenkompetenz). In diesem Zusammenhang – was den Aufbau einer eigenen Datenbasis betrifft – sei Österreich nicht konkurrenzfähig [Interview CWTS]. Für unsere konkrete Aufgabe heißt das: *You can't shrink to greatness*. Staat und Agenturen müssen sich großen Fragen stellen und zur Beantwortung auch Analysekapazitäten und Geld einsetzen und somit auch ambitioniertere Aktivitäten setzen. Solche Bemühungen brauchen jedenfalls eine Datenbasis mit Alleinstellungsmerkmal als Gravitationszentrum, um die herum neue Methoden entwickelt werden können. Diese neuen Methoden (*big data, machine learning*) brauchen große Datenmengen, ansonsten werden methodische Entwicklungen zu theoretischen Trockenübungen. Das können Daten der Bundesstatistik und hoffentlich bald aus dem AMDC sein, das kann eine Wiederbelebung der Forschungsförderungsdatenbank sein.

### **Europa: Research Infrastructure for Science and Innovation Policy Studies (RISIS)**

Hierbei handelt es sich um ein großes, 2014 gestartetes EU-Projekt aus Horizon 2020, das mittlerweile mit RISIS2 seine Fortsetzung über weitere vier Jahre gefunden hat. De facto ist es eine verteilte Infrastruktur über zahlreiche gut ausgewiesene universitäre und außeruniversitäre Forschungseinrichtung, die gemeinsam eine Core Facility betreiben. Die drei wichtigsten Arten von Aktivitäten sind (i) nutzer- und kundenorientierte Tätigkeiten und Programme einschließlich Training und Community Building; (ii) ein Service Layer zur Datenintegration und -analyse einschließlich Methodenarbeit vor allem im quantitativen Bereich und (iii) konzeptive Entwicklung, Aufbau, Pflege und Erweiterung von Datenbanken: Das sind derzeit 13 große Datensätze zu Verständnis und Analyse zentraler forschungs- und innovationspolitischer Fragen zu industrieller Innovation in unterschiedlichen Bereichen und Größenklassen, Patenten, Verflechtungsmustern, Bibliometrie, EU-Projekten, akademischer Mobilität und Hochschulsektoren. Mit dem Science and Innovation Policy Evaluations Repository (SIPER) steht eine weltweite Übersicht zu FTI-Evaluierungen zur Verfügung. Weiters spielen Querschnittsthemen wie Ethik oder RRI in RISIS eine wichtige Rolle.

RISIS will auch methodische Neu- und Weiterentwicklungen<sup>25</sup> sowie eine Erweiterung des Verständnisses in der FTI-Politik erreichen: „... *the project goes beyond established quantitative indicators, developing positioning indicators, which take into account critical features of knowledge dynamics i.e. the importance of asymmetries in producers, in places and in themes.*“<sup>26</sup>

*Kurze Einschätzung:* Wichtige Infrastruktur, die derzeit als Großprojekt gefahren wird. Hier geschieht neben dem Aufbau wichtiger Datenbanken auch abgestimmte Methodenarbeit. Mit AIT und Joanneum Research sind zwei österreichische Einrichtungen im Kern dabei, weitere Einrichtungen wie das WIFO spielen bei einzelnen Aktivitäten wie MORE<sup>27</sup> eine Rolle.

### **Global: Die Ankunft von Big Data**

Auch wenn funktionale Zusammenhänge in der FTI-Politik derzeit noch selten – und wenn nur mit großen Schwierigkeiten – mit Methoden wie AI / Big Data Analysis evaluiert werden (können), gibt es bereits einige Gebiete, in denen diese Entwicklung weit vorangeschritten

ist. Namentlich betrifft das die Bibliometrie und die Patentstatistiken, da in beiden Fällen riesige und verlässliche Datensätze (in der Bibliometrie mit ein paar Abstrichen) vorhanden sind. Beispielhaft dafür stehen (Wu, Wang, & Evans, 2019) zur Frage, wie Teamgröße mit Durchbruchinnovationen auf der Basis von zig Millionen Datensätzen korreliert oder die Übersicht über den Forschungsstand, den die Komplexitätsforscher (Fortunato, et al., 2018) geben. Diese Forschungen bringen wichtige Erkenntnisse zu basalen Zusammenhängen im FTI-Bereich (Beitrag zur Theoriebildung), die dann wiederum für kleinräumigere Evaluierungen genutzt werden können.

*Kurze Einschätzung:* Österreich braucht für diese Themen starke Gruppen und stabile Forschungsinfrastrukturen, wie das Beispiel des Complexity Science Hub etwa im Zuge der COVID-19 Krise in Form seiner einzigartigen Produktivität, Relevanz und Vernetztheit zeigt.

### **Datenzugang und Datenverfügbarkeit**

Die Bedeutung von Daten für die Herstellung von Evidenzen muss an dieser Stelle nicht eigens argumentiert werden. Die Legitimationswirkung, die von Bewertungen und Empfehlungen in Evaluationen ausgeht, hängt wesentlich auf den ihr zugrunde liegenden Daten ab: Je besser die Daten, desto stärker die Evidenz, umso glaubhafter Einschätzungen und Handlungsempfehlungen. Die gerade stattfindende COVID-19 Pandemie zeigt sehr eindrucksvoll – und zum Nachteil dieses Landes – die infrastrukturellen Defizite im Datenbereich ... und genereller im Digitalisierungsbereich und im Bereich der Institutionenlandschaft auf: Der Unterschied heißt: „Robert Koch vs. viele Köche“.

Im Wesentlichen gibt es nach Datenherkunft drei verschiedene Typen, die im FTI-Politik-Evaluierungsbereich eine Rolle spielen bzw. künftiges Potential haben:

- **Daten, die im Rahmen von Tätigkeiten anfallen** und die für die Aufrechterhaltung dieser Tätigkeiten gesammelt und zumindest teilweise genutzt werden. Dabei handelt es sich im Wesentlichen um **administrative Daten** von privaten und öffentlichen Akteuren. Im **privaten, kommerziellen Bereich** gehören dazu Daten, die oft auch unter „Big Data“ kategorisiert werden: Telekommunikationsdaten von Mobilfunkbetreibern, Finanztransaktionsdaten von Banken, etc. Im öffentlichen Bereich könnte man Daten aus der ELGA als Big Data bezeichnen. Darüber hinaus gibt es im öffentlichen Bereich eine Vielzahl von weiteren Daten, die im Rahmen der Verwaltungstätigkeit anfallen und die ein entsprechendes Potential für die sozial- und wirtschaftswissenschaftliche Forschung im allgemeinen und FTI-Politik im Speziellen aufweisen. Im Idealfall sind diese Daten als „Register“ auf Basis einer Rechtsgrundlage eingerichtet.
- **Klassische Statistikdaten**, die aufgrund gesetzlicher Normierung und/oder als Serviceleistung für Dritte erhoben werden. Das können auch Registerdaten sein, sofern sie in die Datensätze der statistischen Ämter aufgenommen werden. Im öffentlichen Bereich sind das **nationale Statistikbüros**. Diese Daten werden in der Regel von der Verwaltung sowie politischen Entscheidungsträger\*innen benötigt, um Entscheidungen in der öffentlichen Politik vorzubereiten, zu treffen, zu rechtfertigen und zu evaluieren. Dies betrifft sowohl eigene nationale Erhebungen aber auch die immer wichtiger werdende europäische Ebene, d.h. internationale Panels oder Surveys. Im privaten Bereich sind das **kommerzielle Datenanbieter**, im FTI-

Bereich insbesondere solche, die Unternehmensdaten sowie Bibliometriedaten anbieten.

- **Ad hoc Daten**, die im Rahmen von Evaluierungen erst erhoben werden.

Jeder dieser drei Typen von Daten hat unterschiedliche Relevanzen für FTI-Politik-Evaluierung und ist mit unterschiedlichen Herausforderungen konfrontiert, wenn man diese in Beziehung zur Struktur der FTI-Politikevaluierungen in Österreich setzt.

In Österreich ist die Struktur des Marktes vor allem durch kleinere Aufträge im nationalen Kontext geprägt (Streicher, Polt, & Unger, 2020). Aus den Budgets ergeben sich daher kaum Möglichkeiten, sehr umfassend und längerfristig Daten zu erheben. Es wird vor allem auf ad-hoc Datenerhebungen gesetzt, um aus diesen das unmittelbare Informationsbedürfnis der Auftraggeber\*innen zu erfüllen und die Fragen aus den *terms of reference* beantworten zu können. Aufgrund der budgetären Beschränkungen erscheint es aber durchaus sinnvoll, bereits vorhandene Daten zu nutzen, um die spärlichen Ressourcen in deren Auswertung bzw. einzelne zusätzliche Erhebungen zu investieren. Dies ist aber in Österreich nicht der Fall, obwohl wir hier eine vergleichsweise gute Ausgangslage hätten:

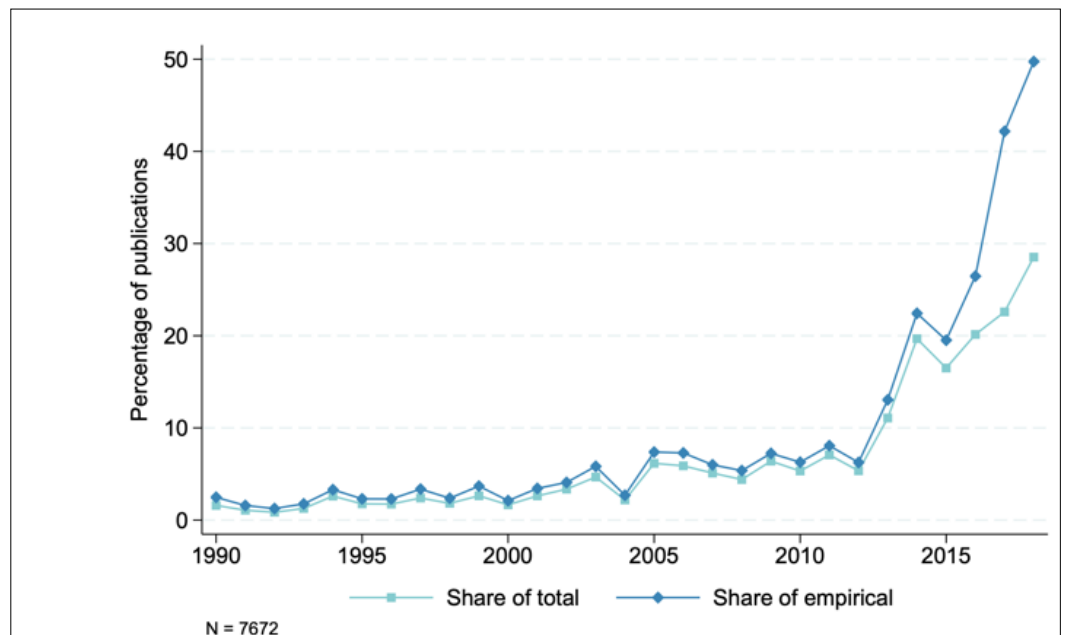
- Es gibt einen umfassenden und für einen Förderstaat relativ stark zentralisierten Verwaltungsapparat, der auf verschiedensten Ebenen Daten sehr umfangreich erhebt – zum Beispiel in Form von **amtlichen Registern**, die in vielen Bereichen der öffentlichen Verwaltung erhoben werden.<sup>28</sup> Einige dieser Daten wären auch für die FTI-Politikevaluierung von besonderer Relevanz, wie das Unternehmensregister bzw. Daten, die staatliche Fördergeber im FTI-Bereich erheben. Im Forschungsorganisationsgesetz (FOG) (Novellierung 2018) wurde eine Rahmengesetzgebung geschaffen, um diese gesetzlich eingerichteten Register für Forschungszwecke nutzen zu können.
- **Statistik Austria** ist mit ca. 700 Mitarbeiter\*innen eine gut ausgestattete Institution, die umfassend und auch gut eingebunden in europäische Kontexte Daten erhebt. Anders als in Deutschland, wo die Statistikämter auf der Länderebene angesiedelt sind, ergeben sich in Österreich durch die Bundeseinheitlichkeit der Erhebungen Synergieeffekte.

Damit enden aber auch schon die Vorzüge und prinzipiellen Vorteile des Standortes beim Datenzugang, und es sei hier auf diverse Defizite und entgangene Chancen hingewiesen:

- Das **Bundesstatistikgesetz** erklärt eine Nutzung der Statistikdaten für die Zwecke der Wissenschaft und Forschung als unzulässig. Die Statistik Austria hat das bisher sehr streng ausgelegt, die Nutzung von Mikrodaten war meist nur über gemeinsame Projekte möglich. Jedenfalls war der Zugang nur mit einigem Aufwand möglich und man ist auf das Wohlwollen der Institution gegenüber dem herangetragenen Vorhaben angewiesen.
- Der Zugang zu **Verwaltungsdaten / Registern** bedarf noch einiger Vorarbeiten, um überhaupt möglich zu sein. Es braucht den konkreten politischen Willen der beteiligten Ministerien (via Verordnung), die Administrativdaten frei zu geben und es bedarf einer technischen Lösung des Zugangs. Dieser technisch einheitliche Zugang wird im Moment unter „Austrian Micro Data Center“ (AMDC) angestrebt. Dieses soll bei der Statistik Austria angesiedelt sein (es gäbe prinzipiell auch andere Optionen), um so einen zentralen Ort des Zugangs zu öffentlichen Daten zu erhalten.



**Abbildung 2:** Nutzung von Administrativdaten in Top-5% Publikationen von wirtschaftswissenschaftlichen Zeitschriften (Halla, M. 2019)



Es gibt hier klare Trends in den Sozialwissenschaften und insbesondere in der Ökonomie. Forschung wird immer empirischer und Nutzung von Administrativdaten steigt. Diesen Trend gilt es auch für die FTI-Politik-Evaluierung zu berücksichtigen.

- Des Weiteren braucht es noch einiger gesetzlicher Anpassung, um die etwas zerklüftete Rechtslage soweit zu vereinheitlichen (Oberhofer, Schwarz, & Strassnig, 2019), dass der Zugang zur Statistikdaten und zu den Registern auf **einer einheitlichen rechtlichen wie technischen Basis erfolgen kann, damit auch eine Verknüpfung der unterschiedlichen Datensätze** möglich ist. Dieser letzte Punkt ist für die Sozialwissenschaften im Allgemeinen aber auch für die FTI-Politik-Evaluierung im Besonderen von großer Bedeutung, da einzelne Administrativdatensätze bereits „ausgeforscht“ sind und die Innovation im Verknüpfen von Datensätzen bestünde. Für die FTI-Politik wäre dies besonders hinsichtlich der Evaluierung von Missionen besonders relevant, da man mit der Verknüpfung von Datensätzen Wirkungen sektorenübergreifend nachgehen könnte.
- Die Nutzung der Daten mit den Haltern von administrativen Registern sowie der Bundesstatistik erfolgt – sofern überhaupt – aufgrund einzelvertraglicher Vereinbarungen. Aus Gesprächen mit Vertreter\*innen der Forschungscommunity kommt hervor, dass die Nutzung der Daten in vielen Fällen möglich ist, aber sehr langwierige Verhandlungen voraussetzen.<sup>29</sup> Die Problematik liegt insbesondere darin, dass dann die Entscheidung über den Zugang bei den Datenhaltern liegt, und damit auch Studien verhindert werden können, die angenommener Weise kritische Befunde produzieren könnten. Im FTI-Politik-Evaluierungsbereich kommt auch eine zeitliche Komponente hinzu: Evaluierungsvorhaben müssen meist unmittelbar begonnen werden, während sich Verhandlungen über den Datenzugang oft ein Jahr oder länger hinziehen können.
- In der überwiegenden Anzahl von in Österreich durchgeführten FTI-Politik-Evaluierungen werden auch eigene Datenerhebungen durchgeführt – oft als Mix von qualitativen und quantitativen Erhebungen (Interviews, Surveys, etc.). Dies **Ad hoc Erhebungen**

werden aber kaum bis gar nicht weiterverwendet. Das liegt auch daran, dass durch die Beauftragungsstruktur den Auftraggeber\*innen die Daten gehören und diese daher nach Auftragsende den Auftraggeber\*innen übergeben oder gelöscht werden. Hier sollte man einen systematischeren Ansatz wählen und (ausgewählte) Datensätze in einem entsprechenden Repositorium archivieren. Dafür braucht es entsprechende Regelungen in den Verträgen sowie die Bereitstellung von entsprechenden Ressourcen (für das Versehen der Daten mit Metadaten, etc.).

- Der größte Budgetanteil in der direkten und indirekten FTI-Förderung (exkl. Uni-Finanzierung) entfällt in Österreich auf den angewandten und daher **unternehmensnahen Bereich**. Daher ist es ein besonderer Nachteil für FTI-Politik-Evaluierungen, dass für diesen Bereich sehr schwer Daten aus öffentlichen Registern und der Bundesstatistik zu bekommen sind – wie etwa die geradezu absurde Situation bei der Evaluierung der Forschungsprämie aufgezeigt hat, dass das BMF als beauftragende Institution entsprechenden Daten (über die sie verfügt) nicht zur Verfügung gestellt hat.
- Es gibt auch einen sehr unterschiedlichen Umgang der **öffentlichen Fördergeber** mit den durch ihre operative Tätigkeit produzierten Daten. Der FWF beispielsweise verfolgt eine offene Datenpolitik. Datensätze können frei genutzt werden. Bei anderen ist die Nutzung nur im Rahmen von Beauftragungen von Evaluierungen möglich, Nutzungen ohne unmittelbaren Anlass durch eine Evaluierung (etwa im Rahmen von akademischen Publikationstätigkeit) sind nicht ohne weiteres möglich. Hier kommen die Datenschutz- und Geheimhaltungsvorbehalte im unternehmensnahen Bereich zum Tragen. Im Rahmen der letzten Regierung bis Mai 2019 gab es Bemühungen, eine Forschungsförderungsdatenbank (FFDB) als Instanz bei der Transparenzdatenbank im BMF zu etablieren. Die Verhandlungen dazu waren auf administrativer Ebene bereits abgeschlossen (im Herbst 2019), es fehlte nur mehr die gesetzliche Umsetzung. Dieses Vorhaben wird in der aktuellen Regierung aber nicht mehr fortgeführt, sodass es im Forschungsförderungsbereich bis auf weiteres keine umfassende Datenbasis geben wird und damit systemischere Betrachtungen bei Evaluierungen erschwert bis unmöglich macht. **Daten für den akademischen Bereich** wären im Prinzip durch die RIS an den meisten Forschungsstätten (insbesondere an den öffentlichen, die diese Daten für die Wissensbilanz benötigen) vorhanden, allerdings finden wir hier eine technisch wie datenmäßig sehr zerklüftete Landschaft vor. Hier gibt es mit ersten Infrastrukturprojekten Bemühungen<sup>30</sup>, gemeinsame Standards und Schnittstellen einzuführen, allerdings erst in einem längerfristigen Zeithorizont.

**Textbox 4:** Internationale Anbieter\*innen am Markt für Forschungsdaten

Wesentliche, für die FTI-Politik Evaluierung relevante Daten sind im hohen Maße proprietär und stehen im Eigentum gewinnorientierter Firmen Dies betrifft insbesondere **Outputdaten wie Publikationen**, die meist im Eigentum der großen **Wissenschaftsverlage** stehen. Der Zugang ist kostenpflichtig, insbesondere wenn Mikrodaten und nicht aggregierte Auswertungen genutzt werden wollen. Sie stellen damit einen relevanten Kostenfaktor bei Evaluierungen da. Neben den großen Verlagen, die zunehmend ihren Datenbestand nutzen, um Geschäftsmodelle im Forschungszyklus vertikal zu integrieren, gibt es auch einige wichtige Anbieter\*innen im universitären Kontext. Hier ist insbesondere das CWTS B.V. – ein Spin-off der Leiden University – zu erwähnen. Dort wurde (unter Nutzung der Web of Science Daten und der weiteren Anreicherung dieser Daten) seit mehr als 20 Jahren ein zentraler bibliometrischer Datenbestand aufgebaut, aus dem auch wesentliche Rankings entspringen bzw. genutzt wird, um im Evaluierungsbereich tätig zu sein.

Seit ca. 10 Jahren sind im akademischen Publikationssektor neben der horizontalen Integration verstärkte Entwicklungen in der vertikalen Integration im Forschungszyklus zu beobachten. Große akademische Verlage wie Wiley, Taylor & Francis und insbesondere Elsevier kaufen entweder Anbieter\*innen von akademischen Services auf oder kommen mit eigenen Dienstleistungen auf den Markt. Der Anteil vertikaler Integration bei Aufkäufen wird insbesondere in den letzten Jahren größer. Die Unternehmensaktivitäten umfassen damit nicht nur den Publikationsprozess selbst, sondern agieren auch zunehmend im Forschungsprozess selbst (George Chen, Posada, & Chan, 2019) (z.B. hat Elsevier 2016 *Hivebench*, ein Laboratory Data Management Tool erworben; Mendeleev, 2013 von Elsevier aufgekauft, bietet neben der klassischen Literaturverwaltung auch ein Datenrepositorium an, in dem man Daten speichern, teilen und veröffentlichen kann).

Für den Bereich der FTI-Politik-Evaluierung ist darüber hinaus sehr relevant, dass die großen Verlage zunehmend auch im Evaluationsbereich tätig werden. Als Beispiel ist hier der Kauf von *Science-Matrix*, einer kanadischen Forschungsevaluationsfirma, im Jahr 2018 zu nennen.<sup>31</sup> 2019 hat Elsevier mit dem *International Center for the Study of Research*<sup>32</sup> eine eigene Organisation gegründet, mit dem Ziel "more effectively evaluate research on dimensions such as productivity, quality, and impact". Das Center – auch unter Mitwirkung hochrangiger AkademikerInnen aus dem Bereich Forschungsevaluation gibt sehr breite Versprechen hinsichtlich ihrer Möglichkeiten ab. Zu den Forschungsthemen gehören „broad impacts of research, and public engagement with research“, Forschungskarrieren in und außerhalb der Forschung, Nachhaltigkeit unter Nennung der UN SDGs. Es werden also Themen genannt, die in der internationalen Diskussion zu FTI-Politik als hochrelevant angesehen werden, wenngleich die bisherige Publikationstätigkeit vor allem auf einer visuell ansprechenden, methodisch aber auf einer deskriptiv-statistischen Auswertung von bibliometrischen Daten beruht, die im Anschein relativ geringen Erkenntniswert für FTI-Politik (Evaluierung) bieten.

Wiley verfolgt eine ähnliche Strategie der vertikalen Integration, allerdings eher fokussiert auf die Ebene der tertiären Ausbildung. Taylor & Francis ist im Vergleich dazu erst sehr spät (2017/18) in den Bereich der akademischen Services eingestiegen. 2017 hat die Gruppe das Start-up colwiz aufgekauft und bietet mit wizdom.ai<sup>33</sup> einen Service an, mit dem mit Hilfe von *artificial intelligence* Evaluierungs- und Foresight Dienstleistungen für Institutionen, Verlage, Fördergeber und Regierungen angeboten werden. Neben der Aktivitäten der großen Verlage, die durch ihren immensen Datenschatz an Publikationsdaten einen großen kompetitiven Vorteil genießen, sind auch eine Reihe von weiteren Unternehmen entstanden, die sich im Feld der FTI (Politik) Evaluierung bewegen oder entsprechende Services entwickeln. Hier ist z.B. *researchfish*<sup>34</sup>, die eine Research Impact Assessment Plattform für Forschungsförderer, Forschungsinstitutionen und ForscherInnen anbieten, zu nennen.

*Dimensions*<sup>35</sup> nutzt neben den klassischen Publikationsdaten auch Altmetric-Daten, Daten zu Grants, Patente und klinische Tests und verlinkt diese Daten untereinander mittels machine learning Techniken. *Dimensions* ist ein Produkt von *Digital Science*<sup>36</sup>, die an der RoRI Initiative beteiligt sind. Inwiefern diese neuen Anbieter\*innen eine Konkurrenz zu den in Österreich vorherrschenden Evaluierungstätigkeiten im FTI-Bereich darstellen, wird sich erst zeigen. Aktivitäten von RoRI z.B. stellen zunächst darauf ab, grundlegende Fragen und basale Zusammenhänge im globalen FTI-System zu erforschen, und bieten damit eine wichtige Ausgangsbasis, auf die dann gezielte Evaluierungsaktivitäten aufsetzen können.

Bei den kommerziellen Anbietern ist die Frage, ob diese mit ihren Daten und Ansätzen komplexere Interventionslogiken von FTI-Maßnahmen überhaupt angemessen erfassen können, da die verwendeten Daten eher eindimensional auf den traditionellen Output abstellen. D.h. für den Moment würden wir diese Aktivitäten eher als ergänzend denn als unmittelbare Konkurrenz zu FTI-Politik-Evaluierung sehen.

*Kurze Einschätzung:* In Österreich ist eine Vielzahl von Daten bei staatlichen Institutionen vorhanden, die für Evaluierungen im FTI-Politikbereich besonders relevant sind. Aufgrund des Partikularinteresses einiger weniger Akteure und eines Desinteresses der Politik an evidenzbasierter Politik bleibt der Zugang und die Nutzung dieser Daten weitgehend und bis jetzt versperrt. Dies zum Nachteil der Forschung, der Transparenz, evidenzbasierter Politik und der Evaluierungskultur. Mit Öffnung dieser Datensätze könnte für die FTI-Politik Evaluierung ein wirklicher Schub zu besserer Evidenz verschafft werden. Ebenso könnten Evaluierungen effizienter durchgeführt werden, weil Daten nicht eigens erhoben werden müssten. Des Weiteren würde das auch die Publizierbarkeit der Evaluierungen in Fachzeitschriften erhöhen, da insbesondere in der Ökonomie auch die Größe und Qualität der Daten durch eine verstärkte empirisch quantitative Ausrichtung zunehmend relevanter wird.

## 3 Befunde aus den Interviews mit Stakeholdern und Expert\*innen

### Zur Befragung

Wie in der Einleitung erwähnt, wurden die Interviews mit den Mitgliedern von FTEval (sowie den Expert\*innen) überwiegend als qualitative Interviews entlang eines Fragebogens geführt. Der Fragebogen bestand aus drei Abschnitten (1) Datengrundlagen, (2) Methoden und (3) Rahmenbedingungen. Jeder Abschnitt hatte Unterpunkte, die mit den Gesprächspartner\*innen im Einzelnen durchgegangen wurden. Diese Punkte beinhalten auch Angaben zu einer quantitativen Einschätzung auf einer fünfstufigen Skala. Zweck der „Quantifizierung“ war, dass die Autoren dieser Studie eine bessere Einschätzung der Relevanz unterschiedlicher Themen erhalten konnten. Der Fokus in den Interviews lag freilich überwiegend auf qualitativen Beobachtungen und Einschätzungen, auch vor dem Hintergrund, dass eine quantitative Bewertung aller Themencluster nicht immer durchgeführt werden konnte. Dies lag zum einen am Typ der Mitgliedorganisation: nicht für alle Mitglieder liegt die vorliegende Studie und die Fragen darin im Kern ihrer Tätigkeiten. Daher konnte in den Interviews nicht zu allen angesprochenen Themen Informationen eingeholt werden – schlicht, weil es für die Organisation keine entsprechenden Anwendungsfälle gab, die im Fragebogen angesprochen wurden. Zum zweiten wurden in den Mitgliedsorganisationen teilweise bewusst Personen um Interviews ersucht, die sonst bei FTEval-Aktivitäten weniger präsent sind, um auch andere Perspektiven und einen breiteren Überblick über die unterschiedlichen Positionen zu bekommen. Drittens ist es gerade bei größeren Organisationen nicht automatisch gegeben, dass es eine Person gibt, die einen vollständigen Überblick darüber hat, welche FTI-politischen Evaluationen durchgeführt werden und aus dieser Sicht spricht, sondern aus der eigenen, evaluationspraktischen Erfahrung heraus.

Das hat zur Folge, dass die quantitativen Bewertungen nicht durchgehend bei allen Fragen eingeholt werden konnten bzw. dass dies oft nicht aus dem Blickwinkel der Gesamtinstitution geschah. Die Bewertungen nach „wichtig“ / „unwichtig“ sind daher als *cum grano salis* zu verstehen und wir nehmen daher davon Abstand, diese auch tatsächlich quantitativ auszuwerten und in dieser Studie darzustellen. Es geht vielmehr um größere Tendenzen (wichtig – weniger wichtig) bzw. wie heterogen die Aussagen zu den Themen waren.

Die Interviews waren von unterschiedlicher Dauer, zwischen 1- 2,5 Stunden pro Interview. Sie wurde entweder einzeln oder mit zwei Gesprächspartner\*innen aus einer Institution geführt. Die Gesprächsinhalte wurden notiert und für die folgende Analyse nach Themenclustern ausgewertet.

Die Autoren bedanken sich bei den Interviewpartner\*innen für ihre ausnahmslos hohe Auskunftsbereitschaft und das gezeigte Interesse am Untersuchungsgegenstand.

### Generelle Reflexionen über die relevante Umwelt der FTI-Politik Evaluierung

Evaluierung und ihr Gegenstand bilden eine Einheit, so dass in den Interviews auch generelle Beobachtungen zum FTI-System geteilt wurden, die wir hier in einer kurzen Punktion wiedergeben wollen:

- Ein starkes, Evaluierungen „auslösendes“ Motiv ist ihre **Legitimationsfunktion**, welche auf Seiten öffentlicher Auftraggeber\*innen besonders stark zum Tragen kommt. Nach der Wahrnehmung einiger Gesprächspartner\*innen ist diese Funktion sogar in Zunahme begriffen [Interviews Auftragnehmer\*innen und öffentliche Verwaltung]. Vor dem Hintergrund, dass eingesetzte öffentliche Mittel vor der Öffentlichkeit und dem Parlament legitimiert werden müssen (*new public management*) kommen nach der Beobachtung von Interviewpartner\*innen auch Aspekte von *message control* [Interview Auftragnehmer\*innen] stärker zum Tragen. Evaluierungen sollen unter einem solchen Blickwinkel dazu beitragen, dass einzelne Maßnahmen erhalten bleiben, sind aber an sich eher ein potenzieller „Störfaktor“ [Interview Auftragnehmer\*innen]. Selbst diejenigen beauftragenden Organisationen, die an sich eher an Learning-Exercises interessiert sind, müssten manchmal in ihren Evaluationen den legitimatorischen Weg einschlagen [Interview Auftragnehmer\*innen]. Durch den fixen Blick auf einzelne Programme und deren Erhaltung gäbe es keinen, der sich darüber Gedanken machte, wie Einzelmaßnahmen in einem Gesamtkonzept zusammenpassten [Interview Auftragnehmer\*innen]. Dieses Set-up hätten dann Effekte auf Terms of References und damit implizite Folgen für die eingesetzten Methoden. Inwieweit tatsächlich die verstärkte legitimatorische Funktion von Evaluierung auch tatsächliche methodische Implikationen nach sich zieht, kann im Rahmen dieser Studie nicht nachgegangen werden.
- In einigen Interviews wird von einem stark **gestiegenem Anspruchsniveau** bei beauftragten Evaluierungen berichtet [Interview Auftragnehmer\*innen]. Bei Aufträgen müsse immer schon alles gekonnt werden und dementsprechend hoch sind die Eintrittsbarrieren. Da spiele die ab und zu fehlende fachliche Kompetenz (über die Möglichkeiten und Grenzen von Methoden) bei der beauftragenden Institution hinein.
- Gleichzeitig wird berichtet und ist auch empirisch beobachtbar, dass **große Evaluierungen seltener** geworden sind, insbesondere summative Ex-Post-Evaluierungen [Interview Auftragnehmer\*innen]. Auch bei (Streicher, Polt, & Unger, 2020; Dinges & Schmidmayer, 2010, S. 258) wird festgehalten, dass Evaluierungen durch kleine Auftragsgrößen und eine formative Perspektive geprägt sind. Insgesamt scheinen begleitende Evaluierungen bzw. das Bringen von Teilevidenzen für laufende Programme ein stärkeres Charakteristikum in der FTI-Politik-Evaluierung zu werden, also in Teilbereichen die Abkehr vom klassischen Politikzyklus. Maßnahmen werden laufend an die Veränderungen in der Umwelt angepasst. International gibt es Hinweise (siehe Abschnitt zu TIPC im Kap. 2), dass ernsthafte Missionsorientierung eine starke begleitende Evaluierung braucht.
- Hinsichtlich der Datenfrage gibt es die Beobachtung, dass in einigen Bereichen Daten in *real time* oder *near time* zur Verfügung stehen, einerseits durch kommerzielle Datenanbieter\*innen (insbesondere bei Publikationsdaten), andererseits durch verbesserte eigene Monitoringdaten, die jederzeit zur Verfügung stehen. Interne Abteilungen dazu werden auf und ausgebaut und Data Scientists dafür eingestellt, um die Informationsbedürfnisse der Institution zu befriedigen [Interview Forschungsförderung]. Vielfach gehe es da auch in die Richtung der Erkennung neuer Trends oder der Erschließung bisher verborgener Kundenkreise.
- Die Bereitschaft und manchmal auch die Möglichkeit, Programme auch enden zu lassen, ist in Österreich recht gering [Interview Auftragnehmer\*innen].

- Die klein strukturierte österreichische Landschaft der FTI-Politikevaluierung sei international kein Wettbewerbsvorteil. Dies betrifft auch die Wettbewerbsfähigkeit auf europäischer Ebene, wo es einen generellen Trend zu größeren Konsortien gibt, die „brutal“ [Interview Auftragnehmer\*innen] für Kleine seien. Aus Sicht der Studienautoren ist dies in einem der *economies of scale* folgenden Feld aber schwer lösbar.
- Die Austragung der FTI-Politik auf dem immergrünen Feld der Programme setzt sich naturgemäß auch in den Evaluierungen fort. Dadurch verfügt die FTI-Politik weitgehend nur über „*Laternenevidenz*“ [Interview Auftragnehmer\*innen], dafür aber über eine Reihe blinder Flecken im Bereich der Institutionen, der Portfolios, beim FTI-Gesamtsystem sowie auf der Meso-Ebene in den thematischen Bereichen [Interviews Auftragnehmer\*innen]. Alles, was bei Maßnahmen und Evaluierung die Grenzen von Verwaltungseinheiten überschreitet, ist politisch und empirisch beinahe unmöglich zu bewältigen.

### **Neue Entwicklungen und Herausforderungen für FTI-Politik-Evaluierung**

Nach neuen Trends und Entwicklungen gefragt, kommen in den Interviews zwei Themen, die das Feld beschäftigten, und zwar nicht sehr überraschend, (1) Wirkungsmessung, vor allem in Richtung Social Impact und (damit auch in einem gewissen Zusammenhang stehend) (2) Missionsorientierung von FTI sowie (3) Indikatorik.

- Das Feststellen von **gesellschaftlichen (inklusive ökonomischen) Wirkungen** bei (nicht nur, aber auch von FTI-) politischen Maßnahmen ist zu einem wichtigen Element öffentlicher Politik geworden. In der FTI-Politik-Evaluierung gibt es eine Kluft zwischen Anspruch – von Seiten der politischen Administration, die ihrerseits der Politik, dem Parlament und der Öffentlichkeit rechenschaftspflichtig ist – und Wirklichkeit des Sag- und Machbaren. Alle Akteure (Auftragnehmer\*innen wie Auftraggeber\*innen) sind sich bewusst, dass Social Impact Messung eine enorme Herausforderung darstellt. Es geht aus Sicht der Evaluator\*innen insbesondere um Erwartungsmanagement, damit Ansprüche an den gesellschaftlichen Impact und dessen Messung nicht überzogen werden [Interview Auftragnehmer\*innen]. Diese Ansprüche kämen vor allem von politischer Seite. Es gibt weiters fundamentale Probleme in der Abgrenzung von Interventionen, es gibt dazu nicht wirklich Daten und es fehlen Indikatoren bzw. ist unklar, ob es dazu überhaupt evidenzbringende Daten und gute Indikatoren geben kann [Interviews Auftragnehmer\*innen und Auftraggeber\*innen]. Ein Politikbereich, der besonders mit recht kleinteiligen Programmen arbeitet, ist von solchen Problemen besonders betroffen. Es fehlen auch die theoretischen Grundlagen und Modelle für Wirkungen [Interview FAS research sowie Interview öffentliche Verwaltung]. Wirkungen sind gleichermaßen eine empirische wie politische Frage (nämlich die politische Anerkennung, was als mögliche erwünschte und relevante Wirkung gilt) und daher sei die Frage nach den Methoden dafür auf dieser abstrakten Ebene nicht beantwortbar [Interview Auftragnehmer\*innen].
- **Missionsorientierung** von FTI ist spätestens seit dem Mazzucato-Papier (Mazzucato, 2018) auf EU-Ebene ein zentrales Thema und spielt wohl künftig bei Evaluierungsaufträgen in EU-Programmen eine zentrale Rolle. Für Österreich bleibt das angesichts der noch fehlenden FTI-Strategie für die kommenden Jahre

abzuwarten. Eine umfassendere Missionsorientierung ist aber nach Einschätzung der Interviewpartner\*innen [Auftragnehmer\*innen] eher nicht zu erwarten, da dies eine sektorübergreifende Strategie erfordern würde, die in der österreichischen Realpolitik nicht sehr wahrscheinlich sei. Gewisse Elemente einer Missionsorientierung seien aber auch für die österreichischen FTI-Strategie möglich.

In den Interviews [Auftragnehmer\*innen] wird festgehalten, dass Missionsorientierung eine grundsätzliche andere Art von Politik bedeutet, in der es nicht nur um FTI-politische Maßnahmen in Form von Programmen geht. Sie beinhaltet auch Regulierungselemente, die alle gemeinsam am Weg beobachtet und gemanagt werden müssen, um Adaptionen vornehmen und Wechselwirkungen beherrschen zu können. Dies habe enorme Konsequenzen für FTI-Politik-Evaluierung. Hier herrscht aber noch relativ große Unklarheit bei den Stakeholdern, was das für Evaluierungsmethoden bedeuten würde. Klar ist der stärkere Fokus auf begleitende Evaluierung, „*aber man wüsste nicht, was man da monitoren würde*“ [Interview Forschungsförderung]. Bemerkenswert ist auch, dass Missionsorientierung zwar immer mehr an Bedeutung gewinnt, aber vor dem Hintergrund, wie FTI-Politik in Österreich aufgestellt ist, nur schwer umsetzbar ist [Interview öffentliche Verwaltung]. Was das methodisch bedeutet, ist schwer abschätzbar, da dies vor allem davon abhängt, wie die Vorgaben der Missionsorientierung auf den unterschiedlichen Policy-Ebenen politisch umgesetzt werden.

- Im Gegensatz zu „Gesellschaftliche Wirkungen“ und „Missionsorientierung“, die im Rahmen einer offenen Frage angesprochen wurde, wurde die Frage der **Indikatorik** in den Interviews als eigene Kategorie abgefragt und auch bewertet. Die Relevanz des Findens von neuen Indikatoren sowie deren Weiterentwicklung wurde als wichtig bis sehr wichtig eingeschätzt, mit einer steigenden Relevanz in den kommenden Jahren insbesondere von Seiten der Auftragnehmer\*innen. Einige Interviewpartner\*innen geben dem Thema auch kaum aktuelle Relevanz. Das Thema sei hier nur kurz angesprochen (siehe auch oben zum deutschen BMBF / vdi-vde Programm), weil fraglich ist, ob dies auf Ebene der Förderung mittels eines Programms bewältigt werden kann. Das Design von neuen Indikatoren ist noch eine verhältnismäßig leichte Frage. Aber, so Aussagen aus den Interviews...
  - Indikatoren sind eine *Frage der Daten* und wie sie erhoben werden können. Hier kommen viele praktische (u.a. finanzielle) und politische Erwägungen ins Spiel. Bei den SDGs wurden die Bemühungen, Indikatoren zu finden, nicht fortgesetzt, weil man den Erhebungsaufwand geringhalten wollte. Die Daten seien nicht da gewesen und es hat sich daher keiner drüber getraut. Manche Sachverhalte seien auch unmessbar [Interview Auftragnehmer\*innen]. An dieser Stelle soll auch erwähnt sein, dass die 2019 sehr weit gediehenen Überlegungen zur Forschungsförderungsdatenbank durch die jetzige Regierung nicht fortgesetzt wurden und daher selbst basale Inputindikatoren de facto nicht systemweit verfügbar sind.
  - Dass Indikatoren zentral eine *politische Frage* sind, wurde in einigen Interviews [Auftragnehmer\*innen] festgehalten. Angesprochen wurde auch die Balance zwischen zu trivialen Indikatoren bzw. auf der anderen Seite und Komplexität bzw. fehlender Nachvollziehbarkeit auf der anderen Seite. Dazu kommt der systemische Charakter von Indikatoren und ihr Zusammenspiel in einem System. Das einfache Hinzufügen neuer Indikatoren kann problematisch sein. Es



brauche einen politischen Prozess, sonst entstünden Datenfriedhöfe [Interview Auftragnehmer\*innen].

- Dort wo es Indikatoren geben würde, weil auch die Daten da wären, werden sie oft nicht genutzt [Interview Auftragnehmer\*innen].
- Für Österreich praktisch ist im Übrigen, dass es das deutsche Programm mit letztlich öffentlich zugänglichen Ergebnissen zu neuen Indikatoren gibt.<sup>37</sup>

Der Vollständigkeit halber seien noch Themen aufgelistet, die von einzelnen Interviewpartner\*innen als relevant in der FTI-Politik-Evaluierung eingeschätzt, aber nicht ausführlich diskutiert wurden: Karriereanalysen (insbesondere von jenen, die aus der Forschung weggehen, Interview Forschungsförderung), zu seltene institutionelle Evaluierungen, fehlende Portfolio-Evaluierungen, sehr seltene Systemevaluierungen, fehlende Evaluierung auf einer Meso-Ebene bzw. von sektorenübergreifenden Themen, Evaluierungen von und an Universitäten [Interview Auftragnehmer\*innen].

### **Relevanz unterschiedlicher Daten für die FTI-Politik-Evaluierung**

Eine entscheidende Frage, die für die Aufgabenstellung für die Studie zentral ist, ist diejenige nach den Daten: Die Möglichkeiten und Grenzen der Datengenerierung und des -zugangs haben wesentliche Implikationen für jede Form von Methodendiskussion in einem bestimmten Feld. Oder anders ausgedrückt: Wenn für Evaluierungen nur Monitoringdaten und selbst erhobene qualitative Daten zur Verfügung stehen, kann man sich beinahe jede weiterführende Methodendiskussion weitgehend ersparen.

Wesentliche Elemente, die wir im Folgenden analysieren, sind: (1) in einem gegebenen Kontext: was sind vorherrschende Strukturen der Datennutzung? (2) Welche Daten eignen sich prinzipiell dafür, um FTI-Policy Fragen zu beantworten? (3) Wie sieht es mit dem tatsächlichen Zugang zu diesen Daten aus?

- (1) Vorherrschend für Beauftragungen in Österreich sind Evaluierungen einzelner Programme und daraus resultierender Fragestellungen nach Impact und Effektivität des einzelnen Programms und den Limitationen aus den Auftragsgrößen (Streicher, Polt, & Unger, 2020). Dies impliziert aus Datensicht vor allem die Nutzung von Monitoringdaten aus den Programmen ergänzt durch Erhebungen im Rahmen der Beauftragungen. Das sind vor allem Interviews und weitere qualitative Methoden sowie Surveys in der relevanten Population. Für die Verortung des Programms in einem weiteren Kontext im Ökosystem von Maßnahmen bzw. in einem globalen FTI-Kontext sind weitgehend kaum Ressourcen vorhanden. Dieser Fokus auf Einzelprogrammevaluation war auch deutlicher Gegenstand von Kritik in den von uns durchgeführten Interviews [v.a. Interviews Auftragnehmer\*innen]. Er ist eine wesentliche Determinante für die zum Einsatz gebrachten Methoden in der Evaluierung.
- (2) Die Datenkategorien, die für FTI-Politik-Evaluierung zur Verfügung stehen, sind – insbesondere hinsichtlich Outcomes/Impacts – auf sehr wenige limitiert. Im Wesentlichen sind es Output-Daten wie Publikationen und Patente. Aus deren Analyse lassen sich kaum Schlussfolgerungen für gesellschaftliche Wirkungen von FTI-Politik herausholen. Die Nutzung anderer Daten dafür ist im Wesentlichen unerprobt und wirft ernsthafte Fragen bezüglich (kausaler) Wirkungen auf. Das gilt vor allem dann, wenn man vor dem Hintergrund einer Programmlogik die Fragen

stellt, d.h. wenn Programm-Owner wissen wollen, was ihr Programm auch gesellschaftlich bewirkt. Gerade vor dem Hintergrund einer aufkommenden Missionsorientierung und steigendem Legitimationsbedarf staatlichen Handelns ist die Erweiterung auf andere Datensets- und -kategorien Voraussetzung, um Veränderungen durch politische Interventionen beobachten zu können (auch wenn aus Gründen der kausalen Zurechenbarkeit die Veränderungen nicht mehr an einer einzelnen Intervention festgemacht werden können).

- (3) Die Misere in Österreich beim Datenzugang, die für FTI-Politik-Evaluierung relevant sind (oder sein konnten), wurde schon weiter oben erörtert. Aufgrund gesetzlicher Vorgaben und weiterer struktureller Mängel wird nur einen Bruchteil der Daten, die eigentlich da wären, genutzt. Und falls die Daten genutzt werden, stecken dahinter oft langwierige Verhandlungen über den Zugang auf Einzelvertragsbasis. (Schwarz, König, Oberhofer, & Strassnig, 2020). In der Öffnung dieser Datensätze liegt ein großes Potenzial für eine methodische Weiterentwicklung – in einem koevolutionären Prozess: Zugang zu den Daten löst methodische Veränderungen aus, und impliziert so, dass weitere Daten für Fragestellungen inkludiert werden können. Es ergibt aber keinen Sinn, methodische Trockenübungen vorzunehmen, d.h. Methoden für diese Datensätze auf abstrakte Weise zu entwickeln, aber tatsächlich keinen Zugang dazu zu haben. Hier ist auch wichtig festzuhalten, wie kritisch die Situation allein in der Frage von tatsächlich erhobenen Daten ist. Das ist noch keine Aussage darüber, welche Daten in Österreich darüber hinaus erhoben werden sollten.

Dieser letzte Punkt spiegelt sich auch in den Interviews wider. Der Zugang zu **administrativen Daten** („Registerdaten“) bzw. Zugang zu den **Daten der Statistik Austria** wird von den Befragten - Auftragnehmer\*innen wie Auftraggeber\*innen einhellig als entscheidender Faktor bei FTI-Politik-Evaluierungen angegeben. Der Zugang zu diesen Daten wäre schon aktuell entscheidend, mit steigender Wichtigkeit in den kommenden Jahren. Die extreme hohe Relevanz wird konterkariert, dass die meisten Befragten diese Möglichkeiten als kaum realisiert ansehen. Es wurde schon an anderer Stelle in diesem Bericht (Kapitel Datenzugang und Datenverfügbarkeit, S. 30) ausführlich über die Situation berichtet. Die Interviewpartner\*innen sehen darin *die* große Leerstelle bei der Politikformulierung, -analyse und -evaluierung in Österreich. Falls es in der Vergangenheit Kooperationen mit Statistik Austria gab, die über für FTI-Politikevaluierung wesentliche Datenbestände verfügt (z.B. F&E Erhebung, ökonomische Daten), waren diese Erfahrungen nicht immer positiv, sodass Zusammenarbeiten auch aufgegeben wurden [Interview Forschungsförderung]. Der Wunsch geht nach einem Zugang auf Mikrodatenebene (also nicht aggregierte Daten), um diese vor dem Hintergrund eines offenen Erkenntnisinteresses explorieren und methodische Innovationen ausprobieren zu können.

**Surveys und Panels** gehören für den FTI-Politik-Evaluierungsbereich aktuell zu einer „sehr wichtigen“ Datenquelle – mit erwarteter leicht steigender Bedeutung über die nächsten Jahre, v.a. aus der Sicht der Auftragnehmer\*innen. Wesentliche Herausforderungen dabei sind aber die Kosten (etwa Umfragen im Firmenbereich), die kaum leistbar seien. Manche Evaluator\*innen sind aus methodischen Gründen sehr zurückhaltend, Surveys aufgrund von Verzerrungseffekten einzusetzen [Interview Auftragnehmer\*innen]. Die Reichweite von Surveys sei auch begrenzt, durch ihren ad hoc Charakter gehen sie über

einzelne Evaluierungen nicht hinaus. Over- bzw. underreporting sind bei Umfragen ein bekanntes Phänomen, daher sollte verstärkt auf Registerdaten gesetzt werden (Angel, Heuberger, & Lamei, 2018). Gerade bei Unternehmen bestehe die Gefahr des Survey Fatigue, insbesondere vor dem Hintergrund, dass Statistik Austria mit der F&E Erhebung bereits viele relevante Daten erhebt und bunkert: Aufgrund des Statistikgesetzes sind diese Daten nicht oder nur mit erheblichem Aufwand zugänglich. Das steigert die Kosten aufgrund von Doppelerhebungen und senkt gleichzeitig die Qualität der Umfragen aufgrund von Survey Fatigue. Insbesondere im FTI-Bereich könnten durch Abstimmung von Statistik Austria mit FTI-Stakeholdern bei Umfragen (insbesondere F&E Erhebung) erhebliche Synergien gehoben werden. Die frei werden Ressourcen könnten auf Seiten der Evaluierenden in Methodenentwicklung und Auswertung gesteckt werden.

Auch bei **Monitoringdaten** wird eine verstärkte Abstimmung mit den Daten-Ownern gewünscht, um schon bei der Erhebung etwaige Evaluierungsvorhaben mitberücksichtigen zu können [Interview Auftragnehmer\*innen]. Bei diesen Daten, die eine hohe Bedeutung für die FTI-Politik-Evaluierung haben – bilden sie doch die Basis für jede Form von Programmevaluierung – wird als Problem identifiziert, dass diese nach Ende des Auftrages gelöscht bzw. den Auftraggeber\*innen übergeben werden müssen [Interview Auftragnehmer\*innen]. Der Nutzen dieser Daten bleibe auf den unmittelbaren Auftrag beschränkt und Verknüpfungen mit anderen Datensätzen sei nicht möglich, wenn man diese etwa für eigene Publikationen nutzen möchte. Da diese Publikationen oft aus akademischem Interesse als Eigenleistung erbracht werden, entgeht der FTI-Politik die Möglichkeit, ohne weitere Kosten etwas lernen zu können bzw. Impact zu erzeugen. Von Vertreter\*innen von FTEval wurde auch beklagt, dass die Forschungsförderungsdatenbank FFDB für die aktuelle Regierung keine Priorität genießt und nicht angegangen wird. Damit ist das FTI-Politik-Evaluierungsfeld mit einer extrem zerklüfteten Landschaft an Datensätzen unterschiedlicher Qualität konfrontiert, deren Zusammenführung eine rechtlich wie ressourcenmäßig quasi unmögliche Aufgabe darstellt. Hier wurde in einem Interview [Auftragnehmer\*innen] festgehalten, dass die FFDB sehr wichtig für Evaluierungen wäre und darüber hinaus auch die Daten der Forschungsstätten miteinzubeziehen wären.<sup>38</sup>

Dies führt auch zur Frage einer **gemeinsamen Dateninfrastruktur**. Diese würde von den Auftragnehmer\*innen als „wichtig“ angesehen werden (sofern es eine solche Bewertung abgeben würde), mit stark steigender Bedeutung einer solchen in den kommenden Jahren. Hier gibt es aber eine ganze Reihe von Vorbehalten, die in den Interviews geäußert wurden.

- Für Daten, die aufgrund eines Auftrages in Evaluierungen erhoben werden, wird von einigen Auftragnehmer\*innen stark dafür optiert, diese in ein gemeinsames **Daten-Repository** (etwa das FTEval-Repository) einzubringen. Eine Sorge ist hier, dass durch ständige Neuerhebungen des bereits Erhobenen auch damit ein Beitrag zum Survey Fatigue geleistet wird. Hindernisse für die Speicherung in einem Repository sind die vertragsrechtlichen Vorgaben bei Aufträgen, die eine Löschung/Übergabe der Daten nach Projektende vorsehen. Dabei geht es ja nicht nur um übergebene Rohdaten, sondern auch solche, die eventuell bereinigt, kategorisiert und weiter angereichert wurden. Zum zweiten müssten die Daten in einem gemeinsamen Repository entsprechend kuratiert werden, damit eine weitere Nutzung möglich ist. Dies ist mit Aufwand und letztlich einem erhöhten Budget bei Beauftragungen verbunden. Drittens gibt es auch Einwände gegen ein gemeinsa-

mes Datenrepositorium aufgrund der Tatsache, dass Evaluierungen durchführende Organisationen auch im privaten Unternehmenssektor angesiedelt sind, und in Bezug auf die Offenheit von Daten in einem kompetitiven Markt auf die eigenen Interessen achten muss (Free Rider Problematik). Vorbildcharakter haben hier jedenfalls die Data Policy des FWF, die verpflichtend einen Data Management Plan im Rahmen von Evaluierungen vorsehen.

- Auch die Frage der **Erhebung gemeinsamer Daten**, die den Mitgliedern von FTEval zugänglich sind, wurde in den Interviews von den Auftragnehmer\*innen angesprochen: D.h. es könnten entweder eigene Surveys durchgeführt oder in Evaluierungen auch Daten nach gemeinsam beschlossenen Vorgaben erhoben werden. Dies würde „eine gemeinsame Strategie voraussetzen und ein Nachdenken, wie man einen gemeinsamen Wissensstand aufbaut, was uns über Projektfragestellungen hinaus interessiert“ [Interview Auftraggeber\*innen]. Die Abstimmung und Konsensbildung darüber sind aber gleichzeitig auch ein Hindernis für eine solche Infrastruktur. Es wurde auch festgehalten, dass es für eine solche Infrastruktur einen neutralen Ort brauche [Interview Auftragnehmer\*innen]. Weiters komme hinzu, dass ein auf die Community beschränkter Zugang Barrieren im Wettbewerb schaffen würde. Hier wurde das Beispiel Deutschland genannt, wo auch FTEval Mitglieder immer wieder Evaluierungen durchführen. Dort sind die Daten nicht auf eine Community beschränkt und erlauben so, auch als österreichische Anbieter\*innen am deutschen Evaluierungsmarkt mitzumischen.

Insgesamt eröffnet die Frage einer gemeinsamen Dateninfrastruktur eine Reihe von Möglichkeiten, die Herausforderungen dabei sind allerdings auch sehr ernst zu nehmen. Es bedarf weiterer Diskussionen und Antworten, die im Rahmen dieser Studie nicht gegeben werden können.

Die Nutzung **kommerzieller Daten** für den FTI-Politik-Bereich wird als „einigermaßen wichtig“ angesehen. Interessanterweise wird hier keine steigende Bedeutung dieser Daten gesehen, trotz der Tatsache, die die Owner der Daten – etwa die großen Wissenschaftsverlage – auch versuchen, in das Geschäftsfeld der Evaluierung einzusteigen (siehe oben Kapitel Datenzugang und Datenverfügbarkeit). Daten aus Web of Science (über den Service Incites) oder Scopus werden genutzt, wichtig sei aber „die Abhängigkeit davon zu senken“ [Interview Auftragnehmer\*innen]. In Ermangelung einer nationalen FFDB, die auch Outputdaten wie Publikationen enthält, ist man insbesondere im grundlagenforschungsnahen Bereich auf diese Anbieter\*innen (auch das CWTS dazu zählend) als *obligatory passage point* angewiesen. Im unternehmensnahen-Bereich der FTI-Politik-Evaluierung spielen die bibliometrischen Daten kaum eine Rolle. Auch hier greift man aber aufgrund des versperrten Zugangs zu Registerdaten auf kommerzielle Daten (z.B. Bureau van Dijk) zu. Diese Datensätze seien aber aufgrund der Verzerrungen darin auch nicht ohne Probleme [Interview Auftragnehmer\*innen].

Zu guter Letzt möchten wir uns noch im FTI-Politik-Evaluierungsbereich einer weitgehend unerschlossenen Datenquelle widmen, nämlich **Big Data**. Ohne uns hier auf die Problematik einer exakten Definition einlassen zu wollen oder zu können, sind darin im Wesentlichen große, unstrukturierte Datensätze gemeint, die oft operativen Tätigkeiten entstammen (d.h. nicht explizit geschaffen wurden, um für etwa wissenschaftliche Zwecke ausgewertet zu werden) und nicht mehr mit herkömmlichen Methoden (klassische Statistikprogramme) ausgewertet werden können. Instruktive Beispiele dafür sind Daten aus dem Bereich Social Media, Zahlungsverkehrsdaten sowie Kommunikationsdaten (etwa Mobilfunk

oder Internet). Auch Daten aus Publikationsdatenbanken kann man dazuzählen, sofern auch die Inhalte der Publikationen (die schwach strukturiert und komplex sind) in einer Auswertung mit aufgenommen werden.

Dieser Bereich wird im Moment von den Interviewten (sowohl Auftragnehmer\*innen und Auftraggeber\*innen) als weitgehend „unwichtig“ eingeschätzt, mit der Erwartung steigender Wichtigkeit in den kommenden Jahren. Es wird erwartet, dass diese Daten künftig eine größere bis große Rolle spielen könnten, allerdings ist den Interviewten kaum eine heimische Studie bekannt, die aus der Perspektive der FTI-Politik-Evaluierung damit gearbeitet hat. Außerhalb dieses Bereiches gibt es aber schon eine Vielzahl relevanter und interessanter Anwendungen. FTI-politische Anwendungsfelder von Big Data werden vor allem bei Trendanalysen in Bezug emergierender Forschungsfelder gesehen oder auch um Kollaborationsmuster zu erkennen (siehe z.B. (Krenn & Zeilinger, 2020; Thurner, Liu, Klimek, & Cheong, 2020) als Beispiele für Publikationen aus Österreich) und auch bei Markt- und Umfeldanalysen [Interview Auftragnehmer\*innen]. Ebenso könne Big Data eine Rolle bei der Evaluierung von missionsorientierten Programmen haben, wo Effekte mit den Standardindikatoren nicht mehr erfassbar sind. Gleichzeitig wurden auch Bedenken geäußert, ob der Einsatz solcher Daten auf Programmebene überhaupt sinnvoll sei. Es stellt sich die Frage der Interpretierbarkeit der Daten vor dem Hintergrund, dass solche Daten auch eine entsprechende Theoretisierung und Modellbildung benötigen (siehe auch das folgende Kapitel „Methoden in der FTI-Politik-Evaluierung“ wo ähnliches zu Netzwerkanalysen gesagt wird). Die Datensätze sind in vielen Fällen intransparent, d.h. es gibt kein Wissen über die Qualität der Einträge oder warum sie im Datensatz verzeichnet sind [Interview Auftragnehmer\*innen]. Zu guter Letzt besteht bei den Ergebnissen Unsicherheit, ob es sich um bloße Korrelationen oder tatsächliche zusammenhängende Wirkungen handelt [Interview Auftragnehmer\*innen]. Die Kompetenzen für eine Auswertung von Big Data sind im Moment vor allem im Informatik- und Physikbereich gegeben, die mit den dortigen Methoden, Programmen und Infrastrukturen diese Daten auswerten können.

### **Methoden in der FTI-Politik-Evaluierung**

Die Frage nach den Methoden steht im Zentrum des Analyseauftrags und -designs.

Wir möchten mit der interessanten, in einem Interview geäußerten Hypothese beginnen, dass Evaluierungsmethoden nicht der Engpass für die Policy-Probleme seien [Interview Auftragnehmer\*innen]. Viel bedeutender seien Probleme in der politischen Abstimmung (zwischen Policy-Ebenen, zwischen gleichrangigen Verwaltungseinheiten) oder die fehlende Aggregation des Wissens hin zu gemeinsam geteilter Evidenz und Theorien über Wirkungszusammenhänge im FTI-Bereich.

Aus Sicht von Evaluator\*innen sind methodische Innovationen eine treibende Kraft in der individuellen wie institutionellen Weiterentwicklung und in der Etablierung von Alleinstellungsmerkmalen am Markt. Ob damit Policy-Probleme und damit des Feldes FTI-Politik-Evaluierung substanziell gelöst werden können, sei allerdings zweifelhaft, wenn das System schon vielfach daran scheiterte, das offensichtlich Erkennbare in entsprechende FTI-politische Maßnahmen umzusetzen. Neue FTI-Maßnahmen sind aber tatsächlich geeignet, Impulsgeber für neue Methoden and Ansätze in der Evaluierung zu werden.

Die These hat etwas für sich und in der Zusammenschau mit den Aussagen der Stakeholder im FTI-Politik-Evaluierungsbereich – und der teils sehr expliziten Ablehnung eines eigenen Programms durch die Auftragnehmer\*innen – bringt dies keine erschöpfende Evidenz hervor, ein für diesen Bereich (und auch von der möglichen Teilnehmer\*innenzahl sehr begrenztes) zugeschnittenes Methodenförderprogramm vorzuschlagen.

FTI-Politik-Evaluierung scheint methodisch von deskriptiv-statistischen Auswertungen und dem Einsatz von klassischen qualitativen Methoden wie Dokumentenanalyse, Interviews bzw. Gruppendiskussion dominiert zu sein. Kursorische Evidenz dafür liefert ein Blick auf die Beauftragungen. Es fehlt aber eine detaillierte Analyse und Auswertung der eingesetzten Methoden, diese findet sich auch nicht in der Metaevaluation (Landsteiner, 2015). Für die 2000er Jahre wurde festgehalten, dass die Datensammlung mittels Interviews überrepräsentiert ist, Peer Reviews fehlen weitgehend (Dinges & Schmidmayer, 2010, S. 256) „simple descriptive statistics, document analysis and context analysis build the core of data analysis methods“ (ebd. S. 257). Anspruchsvollere Methoden (Ökonometrie, Kontrollgruppenansätze, Netzwerkanalysen) wurden nur in Ausnahmefällen eingesetzt. Die bereits erwähnte Tendenz zu kleinen Auftragsgrößen hatte auch methodische Konsequenzen, bis hin zu einem improvisatorischen Vorgehen und der Wahl der einfacheren Ansätze und Daten [Interview Auftragnehmer\*innen].

Im Rahmen der Interviews wurde die Relevanz unterschiedlicher methodischer Ansätze in der FTI-Politik-Evaluierung diskutiert. Dies inkludierte gängige, bisher schon vielfach eingesetzte Methoden, aber auch neuere Ansätze, wie sie aktuell kaum bis gar nicht im Feld zu finden sind. Die Auflistung der diskutierten Methoden entstammt dem Auftrag für diese Studie sowie aus Inputs aus dem Startworkshop. Was klar wird, ist dass die Strukturen des Feldes in einem engen Zusammenhang mit der Methodenwahl stehen. Die politische Kultur in der FTI-Politik, die Wahl der FTI-Politik-Maßnahmen, Datenverfügbarkeit, Dringlichkeit der Evaluierung, ihre politische Funktion, Auftragsvolumen und deren rechtliche Grenzen usw. korrelieren mit der gewählten Methodik. Wie es ein Interviewpartner [Auftragnehmer\*innen] formuliert hat: „*Wahl der Methoden = Not + Anspruchsniveau + Frage + Budget + Was will der Auftraggeber?*“ Die angewendeten Methoden hängen auch von den Kompetenzen auf Seiten der Auftraggeber\*innen ab, die die durch diese Methoden produzierte Evidenz entsprechend verstehen und einordnen können müssen, damit diese für ihren Politikkontext nutzbar gemacht werden können. Gerade bei neuen Methoden bringt das eine Reihe von Herausforderungen mit sich, etwa im Bereich von Machine Learning, welches selbst für diejenigen, die damit arbeiten, weitgehend eine *black box* ist.

Eine Vielzahl weiterer Methoden, die in der FTI-Politik Evaluierung auch zum Einsatz kommen (etwa Delphi, kontrollierte Experimente) können hier nicht ausführlich diskutiert werden (obwohl vereinzelt auch in den Interviews angesprochen).

### **Deskriptive Statistik und qualitative sozialwissenschaftliche Methoden**

Die Interviewpartner\*innen sind sich – bei aller Unterschiedlichkeit der Institutionen und auch der Personen und ihrer Portfolios innerhalb der Institution – sehr einig, dass diese beiden Ansätze zu den „*bread and butter*“ Methoden in der FTI-Politik-Evaluierung gehören und auch die kommenden Jahre hinweg bleiben werden. Hier wird auch keine abnehmende Relevanz für die Zukunft berichtet. Diese Methoden werden als „sehr wichtig“ bis „entscheidend“ gesehen. Deskriptive Statistik sei die „Mutter aller Statistiken“ [Interview Auftraggeber\*innen] und insofern essenziell, als man damit einen Datensatz umfassend verstehen kann (als Voraussetzung dafür, wenn etwa avanciertere Methoden eingesetzt werden) [Interview Auftragnehmer\*innen]. Man könne bei entsprechendem Aufwand auch

sehr viel daraus ziehen. Das gleiche gilt für klassische qualitative Methoden, welche als „Kernstück“ bei Evaluierungen bezeichnet werden [Interview Auftragnehmer\*innen]. Diese Ansätze erlauben das Verständnis von FTI-Politik als „soziales System“ und sie werden auch weiterhin bedeutend bleiben, auch wenn manche Interviewpartner\*innen finden, „es sei nicht innovativ genug“. Es gibt nur wenige Bedenken, die dazu vorbracht wurden, etwa wenn es um die Feststellungen von Wirkungen geht, etwa dass Self-Reporting zu Verzerrungseffekten führen könne [Interview Forschungsförderung].

### **Ökonometrische Methoden**

Ökonometrische Methoden, die avanciertere statistische Beweisverfahren zum Einsatz kommen lassen, werden hingegen schon eher in einer Nischenanwendung gesehen. Darunter kann man Methoden subsumieren, die Kausalitätsannahmen machen. Die Methoden genießen aber immer noch hohe Zustimmung und werden als „sehr wichtig“ angesehen, mit zunehmender Relevanz in den kommenden Jahren. Manchmal werden diese Methoden genutzt, um sich einen USP im Evaluierungsgeschäft aufzubauen [Interview Auftragnehmer\*innen], die Akzeptanz sei aber stärker von den Kompetenzen auf Auftragnehmer\*innenseite abhängig. Die Bedeutung hängt eventuell damit zusammen, dass es von Seiten der Auftraggeber\*innen das Bedürfnis gibt, kausale Effekte in den Maßnahmen nachzuweisen, auch vor dem Hintergrund der Legitimationsfunktion von Evaluierung. Nicht in allen Mitgliedsorganisationen sind für den Einsatz ökonometrischer Methoden die fachlichen Kompetenzen da, es werden dann dafür aber Partner\*innen gesucht. Bei der Rekrutierung diesbezüglicher Humanressourcen gäbe es keine Probleme, da einschlägige VWL-Studienrichtungen in Wien breit und gut etabliert sind [Interview Auftragnehmer\*innen]. Mithin sei der Einsatz von ökonometrischen Methoden aber eine Frage des Budgets, da diese hinsichtlich der Daten, deren Aufbereitung und Analyse (z.B. mittels Kontrollgruppen, Counterfactual-Analyse) aufwändiger sind als deskriptive Ansätze.

### **Weitere qualitative sozialwissenschaftliche Methodenansätze**

Dazu gehören etwa ethnographische Ansätze, partizipative Methoden oder auch *action research* (also der Verknüpfung von Intervention und Beobachtung). Die Einschätzung dieser Methoden durch die Interviewpartner\*innen liegt zwischen „nicht sehr wichtig“ und „einigermassen wichtig“, mit einer großen Varianz in den Antworten. Dort, wo sie durchgeführt werden, werden sie als sehr wichtig angesehen. Quantitativ orientierte Evaluator\*innen schätzen die Relevanz dieser Methoden als weniger wichtig an. Für die kommenden Jahre wird eine leicht steigende Relevanz dieser Methoden gesehen. Viele Evaluator\*innen hätten schon damit gearbeitet und es gibt auch teilweise den Wunsch, mehr damit arbeiten zu können. Der Einsatz dieser Methoden ist aber stark themenabhängig (etwa, wenn sensible Dinge thematisiert werden, die man sehr schwer direkt in Interviews ansprechen kann) und bedingt auch größere Budgets und mehr Zeit als Interviews oder Fokusgruppen. In Bezug auf dafür notwendige Humanressourcen wurden keine Probleme berichtet. Diese Methoden können ohnehin nicht *off the shelf* angewendet werden, sondern müssen auf den jeweiligen Evaluierungskontext angepasst werden. Dabei spielt Praxiserfahrung mit diesen Methoden eine große Rolle, diese sich abstrakt anzulernen könnte eher nachteilig sein. Sie spielen dann auch in formativen Evaluierungen eine stärkere Rolle als es in klassischen Programmevaluierungen der Fall ist, wo es zentral um die Erfassung des Outputs und der Outcomes geht. Da das aber die dominante Auftragslage bei Evaluierungen ist, ist es nicht überraschend, dass diese Methoden eine weniger große Rolle spielen als der Mix aus deskriptiver Statistik und qualitativen Auswertungen.

### **Data und Text Mining / Netzwerkanalysen**

In der Relevanz stehen diese Methoden etwa auf derselben Ebene wie die weiteren qualitativen sozialwissenschaftlichen Methoden (zwischen „nicht sehr wichtig“ und „einigermaßen wichtig“). Ihnen wird aber ein relativer großer Relevanzsprung in den nächsten Jahren zugesprochen (zu „sehr wichtig“). Vom methodischen Anspruch her werden diese von einigen Interviewten als „*nicht zu avanciert*“ bzw. als „*keine methodische Herausforderung mehr*“ bezeichnet [Interview Auftragnehmer\*innen]. Oft reiche es, wenn man eine Person mit einer Affinität dazu einstellt, die sich dann den Umgang mit der Software beibringt. Die Werkzeuge dafür sind leicht verfügbar und bei vielen Evaluierungen im Standardrepertoire. Auf der Nachfrageseite bekommen Data und Text Mining eine zunehmende Relevanz, etwa für Themenevaluierungen oder für Themenradars, um kommende Trends in der Forschung zu erkennen [Interview Forschungsförderung]. Hausintern werden diese Tools auch schon bei Forschungsförderern eingesetzt und es ist zu erwarten, dass diese Kompetenzen künftig ausgebaut werden.

Im Rahmen dieser Studie wurden auch zwei Interviews mit Experten aus dem Feld der Data Science und der Netzwerkanalyse geführt, um eine mögliche Anschlussfähigkeit in Bezug auf Methoden für FTI-Politik-Evaluierung zu explorieren. Data Science ist ein disziplinär sehr breit aufgestelltes Feld und beinhaltet folgende Aspekte: (1) wie sammle ich Daten (nach welchen Kriterien? Gibt es einen Bias?), (2) Datenbereinigung: das ist 80% der Arbeit, (3) Datenaufbereitung, (4) Wahl der Methodik, um die Fragestellung zu beantworten, (5) Daten explorieren: Auswahl der Algorithmen, Auswahl der Visualisierung, Modellierung, eventuell Vorhersagen [Interview Universität Wien]. Dafür gäbe es zahlreiche Methoden und die Arbeit bestehe in der Beschäftigung damit, wie man die richtige Methodik, das richtige Interface aufstellen kann. Abseits von klassischen bibliometrischen Fragen gab es allerdings noch keine Beschäftigung mit FTI-politischen Fragestellungen, auch weil es dahingehend noch keine Anfragen / Kooperationen gab [Interview Universität Wien]. Gerade was Wirkungen und Kausalitäten betrifft, dürfe man sich von Data Science keine *magic solution* erwarten.

Was die Verfügbarkeit von Humanressourcen aus den Data Science für das Evaluierungsfeld betrifft, gibt es sehr unterschiedliche Erfahrungen in der Community. Es wird z.T. berichtet, dass es schwierig sei, Datenexpert\*innen zu finden [Interview Auftragnehmer\*innen]. Die Situation könnte sich entspannen, da an der Universität Wien zwei einschlägige Master-Studiengänge dazu beginnen, die in ein paar Jahren auch für einen besseren lokalen Besatz an Personal sorgen werden. Geplant ist eine breite Abdeckung an Bedürfnissen, d.h. auch angewandte Berufsfelder stehen im Visier [Interview Universität Wien].

Bei den größeren Instituten der Mitglieder von FTEval sind entsprechende Data Science Units entweder geplant oder im Aufbau. Aber auch kleinere Organisation haben Personen, die diesem Bereich zuzurechnen sind. Während die Kompetenz teilweise schon vorhanden ist, fehlt es aufgrund der Gestaltung der Aufträge aber noch an konkreten Anschlussmöglichkeiten in Evaluierungsprojekten.

Ein wichtiger Punkt, der im Interview mit dem Netzwerkanalyseexperten stark hervorkam, ist, dass Netzwerkmethoden (oder auch andere Methoden aus dem Bereich Big Data, Data Science) als solche nicht die methodisch-technische Herausforderung darstellen. Bei den Netzwerkanalysen seien die Tools verfügbar, in den letzten 20 Jahren gäbe es so gut wie keine Innovationen, weil sich an der Formensprache nichts geändert hat. Der Mangel besteht in den Modellen, mit denen man an die Daten herangehen kann: „*Eine Netzwerkkarte ohne Modell ist nicht mehr wert als eine Exceltabelle*“ [Interview FAS research]. Der größte



Aufwand geht in die Entwicklung von theoretischen Modellen über Zusammenhänge, mittels derer die Netzwerke sinnhaft interpretiert werden können. In der Kommunikation der Netzwerke geht es darum, wie diese Modelle als Theorien über Zusammenhänge im FTI-Bereich aufgenommen und akzeptiert werden können.

### **Visualisierung**

Wiewohl keine eigene Methode, steht sie stark mit dem Data Science Feld in Zusammenhang (aber nicht ausschließlich). Visualisierung wird durch die Interviewpartner\*innen für FTI-Politik-Evaluierung als „einigermaßen wichtig“ eingeschätzt, auch hier mit steigender Bedeutung in den kommenden Jahren. Hier sind die Aussagen aus der Community sehr unterschiedlich. Einige Auftragnehmer\*innen finden Visualisierung von Ergebnissen sehr wichtig, andere lehnen sie eher ab. Auch auf der Nachfrageseite schaut es unterschiedlich aus: manchmal wurden Visualisierungen explizit nachgefragt, manchmal wurde entsprechende Visualisierungsangebote bei Aufträgen auch abgelehnt [Interview Auftragnehmer\*innen]. Gerade die Umsetzung einer guten Visualisierung brauche aber entsprechende Ressourcen. Die Herausforderung liege in der Konzeptualisierung und im Informationsdesign (wer ist die Zielgruppe und wie kriege ich die Informationen am leichtesten rüber?), nicht in den Fertigkeiten bei der Benutzung der Software. Dabei könne durchaus auf etablierte Formen der Vermittlung gesetzt werden: *„Balkendiagramme sind zwar schön, aber am leichtesten verständlich“* [Interview Universität Wien] Interaktive Visualisierungen werden dabei immer stärker nachgefragt. Insgesamt wird dieser Bereich aber als nicht entscheidend für die Entwicklung der FTI-Politik-Evaluierung im Methodenbereich eingeschätzt.

### **Artificial Intelligence und Machine Learning Methoden**

Diesen Ansätzen wird in den Interviews mit Auftragnehmer\*innen und Auftraggeber\*innen die geringste Relevanz bei den Methoden zugewiesen und sie werden überwiegend als „unwichtig“ eingestuft. Zwar wird gesehen, dass diese Methoden in den nächsten Jahren wichtiger werden, in der Bedeutung kommen sie aber bei weitem nicht an die klassischerweise im FTI-Politik-Evaluierungsbereich eingesetzten Methoden heran. Die Meinungen gehen hier auch stärker auseinander. Einige wenige sehen eine sehr starke Bedeutung für die Zukunft, während der überwiegende Teil der Befragten diese Methoden auch in Zukunft für FTI-Politik-Evaluierung als irrelevant ansieht.

Dies liegt mithin daran, dass es kaum instruktive Beispiele gibt, wo diese Methoden im FTI-Politik-Bereich angewandt werden. Die meisten Interviewpartner\*innen haben keine Kenntnisse, ob und wo diese Methoden schon verwendet wurde. Wenn man doch schon Anwendungsfälle kennt, findet man die Ergebnisse nicht überzeugend und kaum für den praktischen Einsatz geeignet [Interview Forschungsförderung]. Diese Methoden „geistern“ aber in den Diskussionen herum und generieren so Erwartungshaltungen, die eine Auseinandersetzung damit fordern: *„Alle finden es super, aber eigentlich macht es niemand“* [Interview Forschungsförderung]. Nicht nur deswegen nehmen wir hier auch davon Abstand, AI / ML genauer zu definieren und Anwendungsfälle im FTI-Politik-Bereich danach zu klassifizieren. Die tatsächlichen Anwendungsfälle sind zu gering, um hier in die Tiefe gehen zu können. Auch dem interviewten AI-Experten [Interview TU Wien] sind keine Beispiele bekannt, die im Kernbereich von FTI-Politik zu sehen wären.

Mit AI / ML assoziierte Herausforderungen sind jedenfalls die Ansprüche an die Datenmengen, die dafür verarbeitet werden müssen. Diese sind nur für wenige Aspekte der FTI-Politik vorhanden, etwa Publikationsdaten auf globaler Ebene, Projektdaten auf EU-Ebene.

Anwendungsfälle im Rahmen von Programmevaluierungen im nationalen Kontext werden in den Interviews nicht gesehen. Das schließt Anwendungsfälle von AI / ML nicht aus, müssen aber erst – auch im Kontext eines sich transformierenden FTI-Systems – entwickelt werden. In den Interviews konnten aber AI / ML Anwendungsfälle nur andeutungsweise, wenn überhaupt, imaginiert werden. Probleme gibt es laut Interviews auch mit der Feststellung von kausalen Beziehungen sowie in der Erklärung des Interpretationsmodells gegenüber Auftraggeber\*innen [Interview Auftragnehmer\*innen]: man weiß in vielen Fällen, dass es funktioniert, man weiß aber oft nicht warum. Ansätze, AI erklärbar zu machen, stecken noch in den Kinderschuhen der Grundlagenforschung. Gerade im FTI-Politik-Evaluierungsfeld ist es aber wichtig, in der Bewertung zu kommunizieren, warum eine Maßnahme funktioniert/nicht funktioniert hat, um daraus Veränderungen vorzunehmen oder für andere Bereiche zu lernen. Dies ist ein auch vor dem Hintergrund von zunehmender Kritik an automatisierter Entscheidungsfindung in anderen Bereichen wichtiger Faktor.

Aus dem Gutachter\*innen-Bereich gibt es einige Beispiele des Einsatzes von ML. Dabei geht es darum, für eingereichte Papers oder Projekte relevante Gutachter\*innen zu identifizieren. Für Konferenzeinreichungen wird beispielsweise das *Toronto Paper Matching System* eingesetzt.<sup>39</sup> Das *Research Council of Norway* nutzt ML, um Forschungsprojekte auf Basis von Keywords Forschungsgebiete zuzuordnen. Beide Systeme streifen aber damit nicht einmal an Evaluationsfragen. Im Falle von Norwegen seien die Erfahrungen damit noch nicht so gut, vielfach muss händisch nachgearbeitet werden [Interview Forschungsförderung]. In Luxemburg wurde die Eignung von ML für die Nominierung von Gutachter\*innen getestet und mit den Vorschlägen von Expert\*innen verglichen. Es hat sich dabei gezeigt, dass das ML-Tool schlechter abgeschnitten hat [Interview Forschungsförderung]. ML eignet sich somit eher für Klassifikation und dem Auffinden von Ähnlichkeiten, und genau das könne im FTI-Bereich ein Problem darstellen, z.B. Finden von Gutachter\*innen in sehr kleinen Communities, die sehr geschlossen sind, weil damit eine Außenperspektive fehlt. Aber auch in diesem Bereich der Erkennung von Fachgebieten bzw. dem Auffinden von Gutachter\*innen ist die Datenlage nicht bei allen Institutionen so, dass ML ohne weiteres eingesetzt werden könnte. Es würde dafür große Investitionen verlangen bei einem unbekanntem Mehrwert [Interview Forschungsförderung].

Zusammenfassend kann man festhalten, dass die etablierten Methodensets in der deskriptiven Statistik und der qualitativen Ansätze als hinreichend robust und relevant auch für die nächsten Jahre gesehen – insbesondere in der Kopplung mit dem real existierenden FTI-Feld und der Art wie Maßnahmen gesetzt werden. Gleichzeitig muss aber auch versucht werden, methodische Innovationen ins Feld hineinzutragen und damit zu experimentieren, erste Schritte zu gehen und gemeinsame mit Auftraggeber\*innen einen Lernprozess zu beginnen. Unabdingbar dafür ist jedenfalls eine stark verbesserte Datenlage (und da insbesondere der Zugang zu Administrativ- und Bundesstatistikdaten. Diese Daten existieren und können auch innerhalb einer kurzen Zeitpanne wichtige Impulse für das Feld geben) und wohl auch eine stark veränderte FTI-Politik, die schon allein von sich aus andere Ansätze verlangt. Ein eigenständiges Programm für Methoden im Evaluierungsbereich sehen wir auf Basis der in dieser Studie erhobenen Evidenzen als aktuell nicht zweckmäßig an.

### **Rahmenbedingungen bei beauftragten Evaluationen**

Verbesserungen im Methodenbereich bedingen nicht unbedingt Förderprogramme, sondern haben auch mit Rahmenbedingungen zu tun. Aus diesem Grund wurden auch diese

Aspekte in den Interviews angesprochen, um zu erheben, ob es in diesem Bereich Hebel gibt, um Verbesserungen zu erreichen.

Wie mehrfach festgehalten (Dinges & Schmidmayer, 2010, S. 258; Streicher, Polt, & Unger, 2020), ist der FTI-Politik-Evaluierungsbereich mit relativ kleinen bis sehr kleinen **Auftragsgrößen** konfrontiert. Eine Finanzierung über die Grenzkosten hinaus wird als sehr wichtig aber als kaum realisiert angesehen [Interview Auftragnehmer\*innen]. Nicht durchgehend, aber immer wieder wird von einer Finanzierung unter den Grenzkosten berichtet. Die kleineren Volumina hängen aber auch oft mit klein geschnittenen Evaluierungsgegenständen zusammen. In der Umsetzung wird der Aufwand an das angepasst, was Auftraggeber\*innen zahlen können [Interview Auftragnehmer\*innen]. Dies kollidiert aber oft mit dem Wunsch, innovative Methoden zum Einsatz zu bringen: Auftragsgrößen seien „*lächerlich, um innovative Methoden zu entwickeln und umzusetzen*“ [Interview Auftragnehmer\*innen]. Andererseits seien höhere Preise, etwa um Eigenentwicklungen im Methodenbereich zu finanzieren, im derzeitigen System auch nicht drinnen und im Wettbewerb würden die Preise immer noch unter gehen [Interview Auftragnehmer\*innen]. Zusätzlich kommt ab der selten auch nur annähernd erreichten Summe von 100.000 Euro auch das Vergaberecht ins Spiel, das für Auftraggeber\*innen organisatorisch und rechtlich sehr aufwändig ist und damit Evaluierungsvorhaben noch länger dauern würden. Es wird also stark versucht, größere Aufträge zu vermeiden [Interview Forschungsförderung].

Auf der Auftragnehmer\*innenseite wird eine **Koordination** der öffentlichen Auftraggeber\*innen bei Evaluierungen als sehr wichtig angesehen, allerdings als kaum realisiert. Auf Seiten der Auftraggeber\*innen wird dieser Punkt zwar gesehen, aber als weniger wichtig eingestuft und als ambivalent eingeschätzt. Umfassendere Evaluierungen im FTI-Politik-Bereich sind sehr selten und gehen zurück bis ins Jahr 2008 mit der Systemevaluierung (Aiginger, Falk, & Reinstaller, 2009). Eine letzte einschlägige Übung war das OECD Review 2018 (OECD, 2018), allerdings wurden für diesen Bericht im Rahmen des Backgroundreport keine neuen Daten erhoben, sondern nur die vorhandenen Quellen ausgewertet.

Es gibt den starken Wunsch von einigen Auftragnehmer\*innen, verstärkt Evaluierungen auf einer Meso-Ebene zu größeren Themenbereichen durchzuführen, in denen FTI ein Aspekt ist. Das würde eine Koordination der unterschiedlichen Ministerien voraussetzen. Hier wurde das Beispiel Umwelt genannt und auf finnische Vorhaben verwiesen (Valovirta, et al., 2014) [Interview Auftragnehmer\*innen]. Auch ein regelmäßiger Zyklus aus systemischeren Fragen wurde vorgeschlagen. Was eindeutig festzuhalten ist: Eine Masse an Evaluierungen auf Programmebene kann nicht von der Erkenntnis her als Ersatz für größere systemischere Ansätze dienen, wohl auch, weil sie für Fortschritte und die Transformation in der FTI-Politik auf nationaler Ebene weitgehend irrelevant sind und Charakteristika, die eventuell für ein FTI-System nachteilig sind, perpetuieren können. Mit einer Veränderung des Evaluierungsgegenstandes wären auch Methodeninnovationen zu erwarten, insbesondere wenn man sich bei Themenevaluationen in Konsortien mit Evaluierungsexpert\*innen aus anderen Bereichen zusammenfinden müsste.

Auf Seiten der Auftragnehmer\*innen wird die **zeitliche Komponente** von Evaluierungen als ein sehr wichtiger Punkt angegeben, sprich die längere Dauer von Evaluierungen einschließlich des Zeitraums vor dem offiziellen Projektbeginn. In den Gesprächen zeigt sich dann aber, dass dies insbesondere den Beauftragungsprozess und die Erstellung der Terms of Reference betrifft, die „*manchmal so lange dauert wie die Durchführung der Eva-*

*luierung selbst“* [Interview Auftragnehmer\*innen]. Die Vorlaufzeiten bei der Legung von Offerten (sowohl auf nationaler als auch auf EU-Ebene) seien oft zu kurzfristig angesetzt, was es schwer macht, entsprechende Kapazitäten in den Institutionen dafür frei zu machen. Solche Vorgaben und Strukturen haben jedenfalls keine positive Auswirkung auf die eingesetzten Methoden, da keine Zeit da ist, entsprechende Überlegungen anzustellen. Notwendigerweise muss das genommen werden, was da ist und was man kann [Interview Auftragnehmer\*innen].

Ein größerer **inhaltlicher Spielraum** wird von den Auftragnehmer\*innen im Prinzip als sehr gewünscht aber kaum realisiert wahrgenommen. Von Seiten der Auftraggeber\*innen wird dies als ambivalent und weniger relevant gesehen. Es gibt bereits Wahrnehmungen von stark überfrachteten ToR und die Budgets reichen oft kaum aus, um die gestellten Fragen zu beantworten, sodass kein Spielraum für zusätzliche Fragen gesehen wird

Das **Vorhandensein von internen Evaluierungskompetenzen bei den beauftragenden Institutionen** wird von den Auftragnehmer\*innen vielfach als gut wahrgenommen, insbesondere dort, wo es eine Entsprechung der Kompetenzen im Methodenbereich zwischen Beauftragten und Beauftragenden gibt. Auf diese Weise können die präsentierten Ergebnisse auf Seiten der Auftraggeber\*innen besser verarbeitet werden. Eine Frage, die aber auch auf ein möglicherweise geändertes Verhältnis von Evaluator\*innen und beauftragenden Institutionen hindeuten könnte, ist die zunehmende Bedeutung von internen Analyseabteilungen, z.B. bei Förderorganisationen. Durch die verbesserte Lage bei den eigenen Daten (Digitalisierung) und gut verfügbaren externen Datenanbieter\*innen im kommerziellen Bereich, wird hausintern Wissen generiert, bei dem die Grenzen zwischen Foresight, Monitoring und Evaluation verschwimmen. Auch weil Daten zunehmend in Echtzeit zur Verfügung stehen und auch – etwa über Dashboards – ausgewertet werden können, muss nicht mehr auf eine summative Endevaluierung gewartet werden, um Schwächen und Stärken der Interventionen zu erkennen [Interview Forschungsförderung]. Wenn Programme quasi nicht enden, kommt noch eine weitere schwierige Dimension hinzu. Hier wurden in den Interviews von Auftragnehmer\*innen Beobachtungen geteilt, dass Förderorganisationen eher kleinere Beauftragungen ausgeben, die Teil von internen, umfassenderen Wissensgenerierungsaktivitäten sind (etwa kleinere Recherchen) oder dass Evaluierungen nicht mehr beauftragt werden, weil das schon hausintern gekonnt wird. Diese Frage, da nicht wirklich zentral für diese Studie, kann hier nicht weiter erhellt werden, könnte aber interessante Diskussionen für sich künftige ändernde Rollenbilder im Evaluierungsfeld aufwerfen.

### **Humanressourcen und Rekrutierung**

Dieser Aspekt war beim Startworkshop bereits Thema, mit dem Hinweis, dass es schwer sei, z.B. Data Scientists zu bekommen und auch zu halten. Da Methoden Know-How auch über neues Personal in Organisationen gebracht wird, sind wir dieser Frage auch in den Interviews weiter nachgegangen. Es gibt im Feld der Organisationen einige, die bereits Data Scientists als Mitarbeiter\*innen haben. Man findet sie sowohl bei beauftragenden Organisationen (Forschungsförderer) als auch bei Evaluierungen durchführenden Organisationen – und hier bei größeren und bei kleineren. Größere Organisationen können hier natürlich auch Größenvorteile realisieren und kleinere Einheiten dazu aufbauen. Aus den Interviews geht nicht hervor, dass es hier größere strukturelle Hürden geben würde. Abseits neuer, zunehmend relevanter Bereiche wie den Datenwissenschaften wird über keine

besonderen Rekrutierungsprobleme berichtet, da Österreich und Wien gut ausgebaute Sozial- und Wirtschaftswissenschaften haben, die ausreichend gut qualifizierte Absolvent\*innen produzieren. Beispielsweise sei es kein Problem, Ökonom\*innen zu finden, die eine gute Ausbildung bei quantitativen Methoden haben. Wie sich neue Studiengänge im Bereich der Data Science auf die Rekrutierung im Evaluierungsbereich auswirken, bleibt abzuwarten.

Viele der Mitgliedsorganisationen von FTEval vertreten auch einen akademischen Anspruch, der auf dem Feld der wissenschaftlichen Publikationen befriedigt werden will – entweder weil man sich persönlich daran misst oder die Institution als Ganzes daran gemessen wird. Das macht im sehr angewandten Bereich der FTI-Politik-Evaluation, wo es letztlich darum geht, die Erkenntnisinteressen der Auftraggeber\*innen zu bedienen, ein gewisses Spannungsfeld auf. In einigen Interviews wurde von Auftragnehmer\*innen auch der Wunsch geäußert, auf Basis der durchgeführten Evaluierungen stärker akademisch publizieren zu können. Allerdings sei das bei knappen finanziellen und zeitlichen Ressourcen bzw. engen Evaluierungsgegenständen nicht einfach, akademische Publikationen zu realisieren. Von Seiten der Auftraggeber\*innen wird der akademischen Reputation keine große Relevanz zugeschrieben, ebenso wenig ob aus den Evaluierungen entsprechende Publikationen entstehen (Ausnahme ist der FWF, der ein akademisches Papier explizit als erwünscht kommuniziert). Im internationalen Vergleich ist die Frage der akademischen Reputation von Evaluator\*innen in Österreich als geringer einzuschätzen – z.B. zu auch Evaluierungen durchführenden Universitäten oder außeruniversitären Institutionen in Deutschland, Großbritannien und den Niederlanden. Das hat sicherlich auch mit der institutionellen Aufstellung des Feldes in Österreich und damit zu tun, dass Universitäten bei FTI-Politik-Evaluierung keine signifikante Rolle spielen (siehe Kapitel 2).

### ***Sichtweisen auf öffentlich finanzierte Supportmaßnahmen***

Kernstück dieses Auftrags ist es, Vorschläge für mögliche Maßnahmen – finanziert durch die öffentliche Hand – vorzuschlagen, die dazu führen können, dass über eine verbesserte Methodenkompetenz die Wettbewerbsfähigkeit der Evaluierungslandschaft erhalten bleibt. Um die Sichtweisen der FTEval Mitglieder auf so ein mögliches Programm einzufangen, wurde diese Frage auch im Rahmen der Interviews diskutiert. Dabei haben die Autoren die Eigenschaften eines solchen Programms bewusst nicht im Detail vorgegeben, sondern sich auf folgende Grundcharakteristika beschränkt: kompetitives Programm v.a. für Sozialwissenschaften oder noch enger für die Community, (Weiter-)Entwicklung von Evaluierungsmethoden als Gegenstand.

Auch für uns ein wenig überraschend – und damit nehmen wir das Gesamtbild in dieser Frage vorweg – war, dass überwiegend skeptisch auf ein etwaiges Förderprogramm reagiert wurde – sowohl von Auftragnehmer\*innen- als auch von Auftraggeber\*innenseite. Falls es Zustimmung gab, gingen die Meinungen sehr auseinander, worin die Fördermaßnahme bestehen könne. Nur drei der Organisationen, mit denen wir gesprochen haben, befürworteten grundsätzlich ein “Methodenentwicklungsprogramm” als Fördermaßnahme, der überwiegende Teil reagiert skeptisch bis teilweise stark ablehnend.

Was sind die Gründe, die gegen ein Programm ins Treffen geführt werden?

- Einzelne private Anbieter\*innen lehnen Subventionen in Form von Programmförderung grundsätzlich ab.

- Die thematische Enge eines solchen Programms sowie die vergleichsweise geringe Grundgesamtheit möglicher Begünstigter: Aus Sicht der Autoren dieser Studie ist festzuhalten, dass eine Fördermaßnahme für den FTI-Politik-Evaluierungsbereich auf eine sehr kleine Zahl an Einreichenden stoßen würde, weil von den Universitäten wie bisher schon kaum ein gesteigertes Interesse an FTI-Politik im Allgemeinen und Evaluierung im Besonderen nur aufgrund eines Förderprogramms zu erwarten ist. Es ist daher fraglich, ob ein eigenes Förderprogramm in einer angemessenen Aufwand-Nutzen Relation steht.
- Es gibt die Kritik daran, dass in Österreich die FTI-politische Steuerung dominant über eine spezifische Form von Programmen gelebt wird. Diese Kritik wird vielfach auch von Mitgliedern der Plattform vertreten [Interviews Auftragnehmer\*innen]. Ein weiteres Programm steht daher möglicherweise im Widerspruch dazu. Dieser Einwand wurde in einem der Interviews explizit so geäußert.
- Auch davon ausgehend sehen manche ein mögliches Programm im Kontext der breiten FTI-Politik in Österreich, die selbst eines Veränderungsschubes bedürfe (im Rahmen der neuen FTI-Strategie bzw. des FoFinaG). Diese müsse in Richtung transformativer Politik mit längerfristigen FTI-Zielen gehen, die eine Vielzahl von Sektoren und Instrumenten umfasse [Interview Auftragnehmer\*innen und WU Wien]. Für diese bräuchte es eine wirkungsorientierte Begleitforschung, die Experimentierräume für Reflexion und Forschung eröffnet und wo Politikmaßnahmen und Forschung dazu im Ko-Design entstehen [Interview Auftragnehmer\*innen]. Ein Programm mache nur Sinn, wenn es sich an den Bedarfslagen des Innovationsystems orientiere. Blicke alles beim Alten, würde eine Fördermaßnahme im Evaluierungsbereich nur Mitnahmeeffekte produzieren. Der Zweck müsse die Verbesserung von evidenzbasiertem forschungspolitischem Handeln sein, und dabei werde kein Bedarf an einem Programm gesehen. Projektförderung innerhalb eines spezifischen Programmes wird dabei noch einmal explizit von vielen Auftragnehmer\*innen als auch Auftraggeber\*innen abgelehnt.

Dies gesagt, sprechen sich einige durchaus für ein breites angewandtes sozialwissenschaftliches, öffentlich finanziertes Förderprogramm aus, bei dem Mitglieder von FTEval *auch* einreichen könnten. Oder auch für eine größere Offenheit des FWF in Bezug auf Forschung im sozialwissenschaftlichen Methodenbereich. Im Bereich der angewandten sozialwissenschaftlichen Forschung kann man in Österreich durchaus eine Förderlücke identifizieren, während die Förderung von Technologieentwicklung in technisch-naturwissenschaftlichen Bereichen sehr gut abgedeckt ist. Vorgeschlagen wurden hier dem Jubiläumsfonds der Österreichischen Nationalbank ähnliche Projektvorhaben (mit einer veränderten Kostendeckungsstruktur, die besser auf außeruniversitäre Institutionen angepasst ist. Stichwort: Overheads, Interview Auftragnehmer\*innen). Vorgeschlagen wurde auch, ein Programm wie die „Dynamische Qualitätssicherung“ (siehe Textbox 2) wieder aufzugreifen, das aus einem breiten Bündel an Maßnahmen für die Verbesserung der Qualität in den außeruniversitären GSK bestand.<sup>40</sup>

Mit einer öffentlichen Finanzierung könnten auch **gemeinsame Dateninfrastrukturen** gefördert werden und Datenbestände aufgebaut werden, die einen Mehrwert für den Evaluierungsbereich (und darüber hinaus) bringen könnten [Interviews Auftragnehmer\*innen].

Auch die Förderung von **Weiterbildung / Sabbaticals** wurde mehrfach angesprochen und wird insgesamt als „sehr wichtige“ Supportmaßnahme angeführt. Etwa um eine neue Methode zum Lernen oder um eine Publikation zu verfassen [Interviews Auftragnehmer\*innen]. Gleichzeitig wurden auch einige sehr ernste Bedenken ins Feld geführt, nämlich ob es eine Staatsaufgabe sei, in ganz ausgewählten Bereichen Weiterbildungsprogramme anzubieten [Interviews Auftragnehmer\*innen]. Für Weiterbildung zu sorgen sei auch Eigenverantwortung der Institutionen. Ebenso gibt es mit der Bildungskarenz bereits staatliche finanzierte Möglichkeiten der Weiterbildung.

Der Förderung von **Veranstaltungen und Vernetzungen** wird nur geringe Bedeutung zugemessen, eher sogar explizit abgelehnt. FTEval alleine bietet hier ohnehin ein breites Angebot, an zusätzlichen Formaten wird wenig Bedarf gesehen.

In der Gesamtschau über alle Interviews wird der **Verankerung von FTI-Politik im Hochschulbereich** – und damit einhergehend auch Methodenwissen für Evaluierungen in diesem Bereich – nur geringe Relevanz zugesprochen. Einige schätzen dieses Defizit an fachlicher Expertise an den Hochschulen und insbesondere an den Universitäten als problematisch ein, weil damit auch entsprechendes Systemwissen in der Ausbildung nicht angeboten wird [Interviews Auftragnehmer\*innen]. Die Einwände sind eher „kultureller“ Natur denn aus einer Systemsicht fachlich begründet und werden im Desinteresse der dafür verantwortlichen Politik sowie in der Autonomie der Universitäten in der Personalpolitik und in der Auswahl der Themen gesehen. Dass es diese Themen nicht an den Hochschulen breiter gibt, würden die Autoren dieses Berichts aber eher in der historischen Kontingenz und weniger an einer bewussten Strategie einer Universität sehen, die diese Themen aus argumentierbaren Gründen nicht dort haben will. Am Beispiel der COVID-19 Krise zeigt sich aber ganz allgemein, wie fachliche Leerstellen in einem Land stark negative Auswirkungen auf die Gesellschaft und den Umgang mit Krisen haben können, beispielsweise können Fachgebiete wie Public Health und Epidemiologie in Österreich nicht studiert werden.

Auftragnehmer\*innen, die sich in den Interviews dezidiert dazu geäußert haben, sehen – sofern die damit verbundenen Herausforderungen überwunden werden können – eine solche Maßnahme eher positiv. Es könnte auch Kooperationen der FTI-Politik-Evaluierungslandschaft mit dem universitären Bereich in Österreich stärken. Die FFG hätte mit den Stiftungsprofessuren bereits ein entsprechendes Format dafür, das angewandte Fragen mit akademischer Forschung verknüpft.

Die Finanzierung von **Fellowships** wurde auch vereinzelt als Vorschlag für eine Fördermaßnahme von Auftragnehmer\*innen geäußert – und zwar sowohl Incoming, um Lehrer\*innen auf Zeit an die Institution zu bekommen, oder auch Outgoing, um ins Ausland zu gehen und sich dort entsprechende Methodenkompetenzen anzueignen. Hier gibt es – auch abseits des Evaluierungsbereiches – Lücken bei der Finanzierung von kurzfristigen Auslandsaufenthalten im Rahmen von Fellowships.<sup>41</sup>

*Zusammenfassend* kann man festhalten, dass ein eigenes Programm, das weitgehend nur die FTI-Politik-Evaluierungscommunity adressiert, mit guten Argumenten skeptisch bis ablehnend gesehen wird. Dennoch gäbe es eine Reihe von Maßnahmen – die nicht immer auf eine direkte Förderung hinauslaufen, die zu einer Qualitätsverbesserung in der FTI-Politik-Evaluierungslandschaft beitragen kann. Diese Vorschläge wollen wir im Folgenden Abschnitt machen.





## 4 Empfehlungen

Der Auftrag für diese Studie war es, Empfehlungen für Maßnahmen zu geben, die die Methodenkapazitäten im FTI-Politik-Evaluierungsbereich stärken sollen. Dazu haben wir mit Workshops, ausführlichen Interviews mit Auftraggeber\*innen (Policy-Ebene) und Auftragnehmer\*innen sowie der Analyse ausgewählter internationaler Beispiele versucht, ein möglichst klares Bild über Fehlstellungen, damit verbundene Schmerzen und Behinderungen sowie Möglichkeiten einer Therapie zu gewinnen. Ausgehend von den oben im Text festgehaltenen Befunden stellt sich die Situation für den Vorschlag von Maßnahmen folgenderweise dar:

- Internationale Beispiele demonstrieren, dass der Ausgangspunkt für zielführende Aktivitäten häufig in der Kultur und Ausgestaltung der FTI-Politik selbst liegt, also in der Frage, welche Ziele in welchen kooperativen oder kompetitiven Settings mit welcher Konsequenz erreicht werden sollen. Soll etwa Missionsorientierung ernsthaft angestrebt werden, dann braucht es dafür ein entsprechendes Mindset, ein institutionenübergreifendes Vorgehen und daran angepasste Politikinstrumente. Das sind entscheidende Faktoren und Treiber dafür, welche Instrumente und Methoden dann für Evaluierungen zum Einsatz kommen, siehe die Beispiele von TIPC und Vinnova mit entsprechend groß dimensionierten Evaluierungen. Hier sind wir in Österreich in einer stark von (i) der Unternehmenssubvention und (ii) der Programmlogik getriebenen FTI-Politik und den dementsprechenden Evaluierungsansätzen geprägten Situation verblieben. Der Bereich der Hochschulevaluierung ist qua gesetzlicher Regulierung (Hochschul-Qualitätssicherungsgesetz) auf einen beschränkten Kreis von Marktteilnehmer\*innen limitiert (EQAR-Listung) und verknappt so weiter den Markt für institutionelle Evaluierungen. Ergebnis: Kleine, überwiegend legitimatorische Evaluierungsvorhaben in wenigen Bereichen mit oft unzureichender Datenverfügbarkeit.
- Aus den Gesprächen mit den Expert\*innen und Stakeholdern geht nicht hervor, dass die Methodenfrage das Hauptproblem ist, d.h. dass es nicht sofort zu einem *quick fix* im Methodenbereich kommen muss. Probleme bestehen vielmehr im Policy-Bereich insgesamt, ein Phänomen das durch Aspekte einer Innenorientierung (Legitimation vor allem im eigenen Policy-Umfeld) und durch die Betonung traditionellerer Förderansätze mitentstanden zu sein scheint. Die Methoden, bei denen die österreichische Evaluierungs-Community sich ohnehin ausreichend stark aufgestellt sieht, werden auch als relevante, nachgefragte Methoden jetzt und in Zukunft gesehen, es gibt also auch national keine kurzfristig zu schließende Lücke. Umgekehrt fällt es den Interviewten auf beiden Seiten noch relativ schwer, sich für neue Ansätze auch konkrete Anwendungsfälle vorzustellen. Das heißt aber nicht, dass da nicht etwas Großes kommt, sondern eher, dass emergente Faktoren in der Außenwelt sowie die Weiterentwicklung einer holistischen FTI-Politik vielleicht nicht immer im Zentrum der Aufmerksamkeit mancher ihrer Gestalter\*innen stehen.
- Zugleich gibt es bei aller Verschiedenheit der Herkünfte, Arbeitsbedingungen und Meinungen der Evaluierungscommunity recht weit gediehene überwiegende Einsichten. Dazu gehört, dass viele der Institutionen auf Auftragnehmer\*innenseite ein eng geführtes, für kurzfristige Bedürfnislagen der Community geschaffenes Methodenförderungsprogramm, überwiegend kritisch bis explizit ablehnend sehen. Dagegen sprechen die

geringe Größe des Feldes, die noch kleinere des Marktes, die Gefahr von *sunk costs* und erneuten Abhängigkeiten sowie die Nichtbehandlung von als wichtiger wahrgenommenen Probleme. Hier kommt der Ruf nach längerfristigen, planbaren und vor allem größeren Evaluierungsaufträgen, nach echter Kostendeckung, avancierter Nachfrage und neuen Policy-Ansätzen.

- Der Datenzugang zu existierenden Beständen der öffentlichen Hand ist wohl das wichtigste kurzfristig und vergleichsweise kostengünstig zu lösendes Problem für die Evaluierungslandschaft und für die evidenzbasierte sozialwissenschaftliche Forschung insgesamt. Schon allein daraus ließen sich enorme Effizienzgewinne und auch Methodenimpulse generieren. Ein „Trockentraining“ trägt somit weniger zum Methodenfortschritt bei als zugängliche, substanzielle Datensätze, mit denen analytisch gearbeitet und methodisch Neues erprobt werden kann.
- Im Hochschulbereich gibt es Expertise zu FTI-Politik bzw. dafür relevantes Evaluierungswissen nur in Einzelfällen. Damit stößt man auf Grenzen in der Kooperation und bei den zur Verfügung stehenden Humanressourcen bei neueren und vielleicht künftig relevanten Methoden. Die Expertise im Hochschulbereich im klassischerweise eingesetzten Methodenset (qualitative Sozialwissenschaften, Statistik, VWL) ist freilich kein Problem. Insgesamt sind die Wünsche der Community etwas verhaltener, wenn es um die Stärkung der akademischen Kompetenzen im Bereich FTI-Politik und Evaluierung geht. Dessen ungeachtet plädieren die Autoren der Studie aus ihrer Perspektive heraus sehr wohl für einen staatlichen Impuls, mittels dem der Wissenstransfer und die Kooperation zwischen außeruniversitärer und universitärer Forschung im Bereich FTI-Politik und deren Evaluierung gestärkt wird. Vom Aufbau entsprechender Kapazitäten an den Hochschulen können alle enorm profitieren, auch der außeruniversitäre Bereich: durch Kooperationen in Drittmittelprojekten, durch Heranbildung neuer Expert\*innen sowohl im Methodenbereich als auch in der FTI-Politik, massiv durch die Lehre und Ausbildung bzw. Dissertationsprojekten an der Schnittstelle akademischer und angewandter Forschung. Diese Kapazitäten sind als vorwettbewerblich zu verstehen, d.h. nicht als Ausweitung der Konkurrenz, sondern zur Erarbeitung einer gemeinsamen Wissensbasis, um so avanciertere Evaluierungen zu ermöglichen – weil man etwa in einem Dissertationsprojekt sich schon umfassend und ohne Verwertungszwang mit Methoden spielen konnte; oder weil eine sehr grundsätzliche Hypothese im Bereich der FTI-Politik hinreichend erforscht wurde, so dass eine Evaluierung gut darauf aufsetzen kann. Die Autoren sind der Auffassung, dass der Hochschulsektor im Bereich der FTI-Politik gegenüber fast allen Vergleichsländern eine Fehlstelle aufweist: diese zu adressieren und Kapazitäten aufzubauen würde allen zugutekommen.

Die in der Folge gegebenen Empfehlungen versuchen, diesen Befunden Rechnung zu tragen.

### **Argumentation einer staatlichen Interventionslogik**

Die Finanzierung einer spezifischen Supportmaßnahme bedarf auch einer argumentierbaren Logik für eine Intervention. In Bezug auf eine relativ eng gefasste Maßnahme – eine, die im Wesentlichen Akteure am Feld der FTI-Politikevaluierung mit überschaubaren, kompetitiven Projektförderungen adressiert – lassen sich folgende Argumente ins Feld führen:

- Der FTI-Politik-Evaluierungsmarkt in Österreich ist mit etwa 750.000 Euro jährlich sehr klein; ebenso die Zahl der durch eine eng geschnittene Fördermaßnahme

begünstigten Akteure. Eine große Fördermaßnahme ließe sich kaum in Bezug zur Marktgröße rechtfertigen. Bei einem kleinen Programm stellt sich die Frage der Effizienz.

- Die Art der Aufträge in diesem Markt bietet eine ungünstige Ausgangslage für methodisch-analytische Weiterentwicklungen auf Seite der Fördernehmer\*innen, da überwiegend kleinere Aufträge im Rahmen von Programmevaluierungen vergeben werden. Rasche Anwendung, Diffusion und Transmission sind somit fraglich: In diesem FTI-politischen Rahmen ist selbst mit einer entsprechenden Fördermaßnahme kaum Methodeninnovation zu erwarten.
- Für Aufträge auf europäischer Ebene sind in erster Linie Größenvorteile von Relevanz (Frage von größeren Konsortien, Frage von kritischen Massen an Personal und Erfahrung, d.h. größere Institutionen bei den Auftragnehmer\*innen). Hier besteht höchste Unsicherheit wie dieser Herausforderung mittels einer öffentlichen Supportmaßnahme zu begegnen wäre. Wenn hingegen für den internationalen Markt elaboriertes Spezialwissen aufgebaut wird, kann damit lokal mangels Nachfrage nicht häufig gepunktet werden.
- Für kleinere Maßnahmen und Förderungen gibt es schon teilweise staatliche Förderinstrumente (etwa im Bereich der Weiterbildung in Betrieben durch die Bildungskarenz), sodass eine spezialisierte Maßnahme für eine kleine Gruppe an Begünstigten politisch nicht argumentierbar erscheint. Hier liegt die Antwort in einer doch etwas größeren Zahl von Begünstigten, siehe die Empfehlung, angesichts der aktuellen Krise das Programm „Dynamische Qualitätssicherung“ für die außer-universitären Sozialwissenschaften wiederzubeleben.

### **Datenzugang, Datenzugang, Datenzugang!**

Die Datenfrage steht in einem engen Zusammenhang mit den eingesetzten Methoden. Hier lassen sich aus den Interviews und vieler weiterer Evidenzen große Defizite für den österreichischen Kontext feststellen. Mit neuen Daten kommen neue Methoden und neue, nun oft besser beantwortbare Evaluierungsfragen.

Der Zugang zu Daten der Bundesstatistik ist weitgehend versperrt oder nur über wenig befriedende Umweglösungen möglich. Dasselbe gilt für (gesetzlich eingerichtete oder anderweitig eingeführte) Register oder Verwaltungsdatensätze.

- Bemühungen zur Einrichtung eines *Austrian Micro Data Centers AMDC*, mit welchem die Datenbestände der Statistik Austria sowie weiterer Verwaltungsregister zugänglich gemacht sowie untereinander verknüpfbar werden können, werden explizit gefordert. Es braucht dafür verstärkten Druck aus der Community sowie von den Ministerien, in deren Interesse qualitativ hochwertige FTI-Politik-Evaluierungen sind, dass dieses Vorhaben auf gesetzlicher Ebene umgesetzt wird und auch eine entsprechende Ressourcenausstattung für eine professionelle und serviceorientierte Umsetzung erhält. Das Vorhaben ist im aktuellen Regierungsprogramm enthalten, dem voraus gehen aber schon viele Jahre vergeblicher Lobbyarbeit. Experimente mit Big Data etc. sind im Einzelfall hilfreich und instruktiv, was es braucht, sind *bread and butter* Daten, die eine solide empirische Basis für das FTI-Politikfeld bieten. Das AMDC böte eine datenschutzkonforme Zugangsmöglichkeit zu Individualdaten. Ebenso blieben durch die Art des Zugangs und durch die recht-

liche Regelung (Output-Kontrolle, Akkreditierungen von Institutionen und Forscher\*innen, Strafbestimmungen) Geschäftsgeheimnisse auf Unternehmensseite gewahrt.

Ein Spillover eines solchen Datenzugangs ist auch abseits von Evaluierungen insgesamt für die Sozial-, Wirtschafts- und auch Lebenswissenschaften am Standort zu erwarten. Von der dadurch bedingten Qualitätssteigerung durch internationale Wettbewerbsfähigkeit kann wiederum der Evaluierungsbereich profitieren, etwa durch die gemeinsame Arbeit an Datensätzen, methodischem Wissenstransfer und Humankapital.

- *Register und Verwaltungsdatensätze frei geben.* Des Weiteren wird empfohlen, dass sich die an FTEval beteiligten Ministerien dafür einsetzen, um Register und Verwaltungsdaten in ihrem Wirkungsbereich in das AMDC einzuspeisen und so auch für FTI-Politik-Evaluierung einfache und einheitliche Nutzung zu ermöglichen.
- *Einrichtung der Forschungsförderungsdatenbank (FFDB):* Die bereits auf Verwaltungsebene ausverhandelte Lösung ist nicht mehr Teil des laufenden Regierungsübereinkommens und soll nach Aussagen politischer Vertreter\*innen gemäß der ausverhandelten Pläne nicht umgesetzt werden. Die Zukunft steht daher in den Sternen, auf absehbare Zeit wird es keine Lösung geben. Gerade für die FTI-Politik-Evaluierung wäre diese Datenbank (sofern die Datenbank auch Output-Daten enthalten hätte) so etwas wie die *bread and butter* Datenbank gewesen. Dass es hier keine Lösung geben soll, ist für die FTI-Politik und die Evaluierung derselben kein akzeptabler Zustand. Hier sind die Ministerien gefordert, FTEval soll auch auf eine neue Lösung drängen. Eine Unterbringung dieser Daten bei einem etwaigen AMDC ist aufgrund der Synergieeffekte zu bevorzugen und wird dringlich empfohlen: Keine Insellösungen!
- Es gibt auch Möglichkeiten, dass die FTEval-Community im eigenen Bereich um eine verbesserte Datenlage sorgt: durch die Errichtung eines *gemeinsamen Datenrepositoriums* (im FTEval Repository). Hier sind aber noch viele Bedenken innerhalb der Community auszuräumen. Ein wesentlicher Faktor dabei ist, dass die Daten von beauftragten Evaluierungen im Eigentum der Auftraggeber\*innen stehen – d.h. meistens dem Staat. Hier bietet FTEval als Plattform die Möglichkeit, eine entsprechende Policy mit staatlichen Auftraggeber\*innen auszuhandeln.

### **Vorschläge für Maßnahmen auf Auftraggeberseite**

Die Auftraggeber\*innenseite – Ressorts wie Agenturen – können einige sehr hilfreiche Handlungen setzen. Hier gibt es wenig neues zu erfinden, sondern einiges umzusetzen:

- *Strategisch agieren:* Dazu gehört auch, dass ein Umsetzen der Ergebnisse der aws / FFG Evaluierung (Bührer, et al., 2017) die Verantwortung für die Körnigkeit und Taktung von Programmen von den Ressorts stärker hin zu den großen Förderagenturen verlagern würde / wird / kann. Damit kann etwa die FFG ein Portfolio evaluieren lassen oder ein Programm einmal einstellen und neues erproben; sie kann zeitliche Planungssicherheit für alle Beteiligten garantieren und auch von Zeit zu Zeit größere Evaluierungen beauftragen. Beim FWF ist das ja bereits der Fall. Dies findet seine Ergänzung darin, dass sich die Ministerien stärker um missionsorientierte Policy-Portfolios kümmern können.

Vielleicht ist die COVID-19 Krise ein guter Anlass für eine Weiterentwicklung in Richtung umfassenderer Politikansätze, größerer (Programm-)Interventionen und dazugehörigen größeren interim- und ex post-Evaluierungen.

Praktisch alle Auftraggeber\*innen und Evaluators\*innen in Österreich im FTI-Bereich können von einer solchen holistischen und verbindlicheren Politikgestaltung profitieren. Das FoFinaG und die anderen Teile der Forschungsfinanzierungsnovelle 2020 (Pakte und Finanzierungsvereinbarungen) bieten einen guten Rahmen für die Diskussion und eine mögliche Neustrukturierung auch von Evaluierungen, siehe Kap. 2.

- *Größere Aufträge (1)*: Selbst wenn die traditionelle Programmförderung weiterhin das Instrument der Wahl bleibt, sollte es in Österreich weniger kleinstrukturierte Programme geben (wie von OECD oder dem Rat für FTE empfohlen) und einige von ihnen sollten auch ein Enddatum haben dürfen. Damit würde sich die Frage legitimierender Zwischenevaluierungen auf ein geringeres Maß reduzieren lassen. Es würde viel Raum und Ressourcen frei, um über andere Evaluierungen nachzudenken und damit auch über neue Methoden.
- *Größere Aufträge (2)*: Das VINNOVA Beispiel zeigt, wie es gehen kann. Die große staatliche Subventionsagentur muss nicht auch noch ihre Auftragnehmer\*innen subventionieren, sondern erlaubt ihnen im Rahmen von großen Evaluierungen zu experimentieren, zu lernen und das Geld zu verdienen, das ein Dienstleister eben braucht, um seine Produkte auf hohem Niveau anbieten zu können. In Schweden sind die aktuellen Großevaluierungen auch nicht immer vorhanden gewesen, sondern waren Teil eines schmerzhaften Suchprozesses, verbunden mit einem Eingeständnis: *Schon klassische FTI-Politik ist schwer zu heben und zu bewältigen, aber bei großen Missionen hört sich die konventionelle Weisheit endgültig auf.* Daher wurden neue und größere Formate auch bei (begleitenden) Evaluierungen eingeführt.

### **Vorschläge für öffentliche Supportmaßnahmen**

Auch wenn aufgrund starker Argumente eine Supportmaßnahme spezifisch für die österreichische FTI-Evaluierungscommunity nicht empfohlen wird, gibt es in den Sozialwissenschaften, gerade was eine anwendungsorientiertere Ausrichtung betrifft, Wirkungsbereiche, die die Nöte dieses Bereichs lindern können und zu einer gesteigerten Qualitätsorientierung beitragen können. In diesem Rahmen können auch methodische Fragen einen Raum einnehmen.

- Wiederauflegen des Programms *Dynamische Qualitätssicherung*, wie es bereits einmal vor weit mehr als zehn Jahren durchgeführt wurde. Das ist auch angesichts der Corona-Ausfälle ein Instrument, damit die Kleineren bzw. solche, die stark von Aufträgen abhängen, insgesamt überleben können (zu den Details des Programms, siehe Textbox 2 oben). Im Programm wurde eine Vielzahl von Zielen verfolgt, es umfasste Maßnahmen wie Methodenerwerb, Weiterbildung, disziplinen- und institutionenübergreifendes Arbeiten, Anschubfinanzierung für Größeres, Publikationen, etc. Das Programm würde auch anderen Institutionen und Themenfeldern im sozialwissenschaftlichen Bereich offenstehen und wäre letztlich auch ein qualitätsfördernder Ausgleich für den Ausfall vom Geschäft durch die Krise.
- *Größeres Forschungsprogramm für die Sozialwissenschaften im angewandten Bereich*, das Bereiche ins Visier nimmt, die aufgrund der knappen Mittel und der

Ausrichtung beim FWF nicht finanziert werden können. Das heißt: Fokus auf Projektförderung für Fragen, die anwendungsnahe sind und das auch zeigen müssen; neue interdisziplinäre Ansätze und Methoden ausprobieren, wo ein akademischer Pay-off an Publikationsleistungen nicht unmittelbar gegeben ist; Nutzer\*innen des produzierten Wissens von Anfang an mit an Bord nehmen. Der gesellschaftliche Nutzen muss demonstriert werden anhand konkreter Objekte und Themen, die aus der Praxis kommen. Methodenentwicklung am Anwendungsfall würde auch in diesen Programmrahmen passen. Hier wurde schon seit längerem von einigen Akteuren ein Defizit in der Förderlandschaft festgestellt. Die Förderbedingungen müssten dergestalt sein, dass ein solches Programm auch für Institutionen attraktiv ist, die über keinerlei Basisfinanzierung verfügen. Hier sprechen wir von einem Beitrag zu Overheads und Anerkennung tatsächlicher Personalkosten, anders als das etwa aktuell beim Jubiläumsfonds der Österreichischen Nationalbank der Fall ist, deren Fördermaßnahme von einer substanziellen Grundfinanzierung der teilnehmenden Organisation ausgeht. Klarer Weise wäre die Frage der Policy- und Evaluierungsforschung von einem solchen größeren Programm mit angesprochen.

- Ausbau der *akademischen Kapazitäten* in Bereichen, die für FTI-Politik-Evaluierung relevant sind. Dazu gehören der an Universitäten unterrepräsentierte Bereich der FTI-Politik, wie er in praktisch allen anderen Ländern wie Deutschland, den Niederlanden, den skandinavischen Staaten und auch Großbritannien an Universitäten in größeren Einheiten stark vertreten ist. Von dort kommen auch viele methodische Innovationen und auch relevante Grundlagenforschung zur Thematik. Ein wichtiges Schlagwort ist *capacity building*. Das Instrument der Wahl wäre das – um einige Elemente deutlich anzupassende – Instrument der Stiftungsprofessur,<sup>42</sup> wie sie bisher die FFG durchführt. Professuren sollten so eingerichtet werden, dass Kooperationen mit dem außeruniversitären Bereich zu ihrer DNA gehören, etwa über „Double Affiliations“, mit Teilzeit an der außeruniversitären Organisation, Betreuung von Studierenden (Master, Dissertation), Entwicklung eines gemeinsamen Masterprogramms oder einer Doctoral School unter Einbeziehung weiterer Institutionen sowie Inklusion der außeruniversitären Forscher\*innen in die Lehre. Damit sind auch größere Ressourcen durch Pooling möglich, und damit die Attraktion von besonders leistungsstarken Forscher\*innen. Das Beispiel Deutschland zeigt, dass solche Modelle auf vielfältige Weise gut funktionieren können. Gemeinsame Projektanträge für Drittmittelprojekte sind denkbar. Der Sinn ist, dass an den Hochschulen in der Methodenforschung schon auf die Besonderheiten der FTI-Politik-Evaluierung Rücksicht genommen wird bzw. umgekehrt, Absolvent\*innen bereits mit entsprechendem Domain-Knowledge in eine außeruniversitäre Organisation oder bei anderen Akteuren eintreten können. Dieser Bereich muss gerade in den Sozialwissenschaften noch weiterentwickelt werden, die Beispiele sind im Moment nicht sehr zahlreich. Es gilt bei der Ausgestaltung allerdings noch eine Vielzahl von Herausforderungen hinsichtlich Widmung, Auswahl und Finanzierung zu adressieren.

In diesen Maßnahmenbereich fällt auch der Auf- und Ausbau von Forschungs- und Analysekapazitäten an bereits wissenschaftlich starken Institutionen/Kooperationen, die über die Jahre hinweg wichtiger werden, auch wenn sie derzeit noch nicht als drängend angesehen werden (weil sie noch ohne Basisfinanzierung auskommen müssen).

Damit sind jedenfalls stabile finanzielle Arbeitsbedingungen für hochqualitative Arbeitsgruppen und Institutionen im Bereich avancierter Datenanalysen wie Big Data Analysis oder Komplexitätsforschung gemeint, wie sie am Complexity Science Hub durchgeführt werden.

## 5 Literatur

- Aiginger, K., Falk, R., & Reinstaller, A. (2009). Systemevaluierung der österreichischen Forschungsförderung und -finanzierung: Kurzzusammenfassung des Endberichts. Projektbericht. Systemevaluierung der österreichischen Forschungsförderung und -finanzierung. Wien: doi: 10.22163/fteval.2015.55.
- Angel, S., Heuberger, R., & Lamei, N. (2018). Differences Between Household Income from Surveys and Registers and How These Affect the Poverty Headcount: Evidence from the Austrian SILC. *Social Indicators Research*, 575–603.
- Buchinger, B., Marchner, G., & Schoibl, A. (2007). Qualität und Qualitätssicherung in der außeruniversitären gesellschaftsbezogenen Forschung. Grundlagenpapier. [http://www.forschungsnetzwerk.at/downloadpub/Grundlagenpapier\\_gesellschaftsbezogener\\_forschung.pdf](http://www.forschungsnetzwerk.at/downloadpub/Grundlagenpapier_gesellschaftsbezogener_forschung.pdf).
- Bührer, S., Daimer, S., Koschatzky, K., Sheikh, S., Kaufmann, P., & Ruhland, S. (2017). Evaluierung der Förderungsgesellschaften Austria Wirtschaftsservice GmbH (aws) und Forschungsförderungsgesellschaft mbH (FFG). Karlsruhe. doi: 10.22163/fteval.2017.266
- Bundeskanzleramt et al. (2011). Der Weg zum Innovation Leader. Strategie der Bundesregierung für Forschung, Technologie und Innovation. [https://www.bmk.gv.at/dam/jcr:38ab2c5e-5d8f-4502-bbd2-f54f94df9bf1/fti\\_strategie.pdf](https://www.bmk.gv.at/dam/jcr:38ab2c5e-5d8f-4502-bbd2-f54f94df9bf1/fti_strategie.pdf)
- Dinges, M., & Schmidmayer, J. (2010). Country Report: Austria. In J. Edler, P. Cunningham, A. Gök, J. Rigby, E. Amanatidou, I. Garefi, et al. Understanding evaluation of innovation policy in Europe. INNO-Appraisal Final Report (S. 243–264). Manchester: doi: 10.22163/fteval.2010.141.
- Elg, L., & Hakansson, S. (2012). Impacts of Innovation Policy. Lessons from Vinnova's impact studies. Vinnova Analysis VA 2012:01.
- Fortunato, S. et al. (2018). Science of science. *Science* (359, 6379), doi: 10.1126/science.aao0185
- FTB (2016). Österreichischer Forschungs- und Technologiebericht 2016. Lagebericht gem. § 8 (1) FOG über die aus Bundesmitteln geförderte Forschung, Technologie und Innovation in Österreich. Wien: doi:10.22163/fteval.2016.260.
- FTB (2017). Österreichischer Forschungs- und Technologiebericht 2017. Lagebericht gem. § 8 (1) FOG über die aus Bundesmitteln geförderte Forschung, Technologie und Innovation in Österreich. Wien: doi:10.22163/fteval.2017.261.
- FTB (2018). Österreichischer Forschungs- und Technologiebericht 2018 Bericht der Bundesregierung an den Nationalrat gem. § 8 (2) FOG über die Lage und Bedürfnisse von Forschung, Technologie und Innovation in Österreich. Wien. doi: 10.22163/fteval.2018.286.
- FTB (2019). Österreichischer Forschungs- und Technologiebericht 2019. Bericht der Bundesregierung an den Nationalrat gem. § 8 (2) FOG über die Lage und Bedürfnisse von Forschung, Technologie und Innovation in Österreich. Wien. doi:10.22163/fteval.2019.438.
- George Chen, G., Posada, A., & Chan, L. (2019). Vertical Integration in Academic Publishing. Implications for Knowledge Inequality. Connecting the Knowledge Commons — From Projects to Sustainable Infrastructure: The 22nd International Conference on Electronic Publishing – Revised Selected Papers. Marseille: Open Edition Press. doi: 10.4000/books.oep.9068
- Halla, M. (2019): How Did Administrative Data Change Economic Research Output – A Descriptive Analysis. Annual Meeting of the Austrian Economic Association 2020. WU Wien — February 25, 2020.
- Hesse, F., Matt, I., Reckling, F., Völker, T., & Possanner, N. (2019). Standortbestimmung der Bildungsforschung in Österreich. Wien. [https://zenodo.org/record/3566077/files/Bildungsforschung\\_Bericht%20und%20Anhaenge\\_Version2.pdf.pdf?download=1](https://zenodo.org/record/3566077/files/Bildungsforschung_Bericht%20und%20Anhaenge_Version2.pdf.pdf?download=1)



- Hook, D., Porter, S., & Herzog, C. (2018). Building Context for Search and Evaluation. *Frontiers in Research Metrics and Analytics*. doi: 10.3389/frma.2018.00023.
- Krenn, M., & Zeilinger, A. (2020). Predicting Research Trends with Semantic and Neural Networks with an application in Quantum Physics. *PNAS* (117, 4), S. 1910-1916.
- Landsteiner, G. (2015). Nützlichkeit und Nutzen der Programmevaluationen im Bereich der österreichischen FTI-Politik. Metaevaluation der Programmevaluationen 2003–2014. doi: 10.22163/fteval.2016.4.
- Mazzucato, M. (2018). Mission-Oriented Research & Innovation in the European Union. A problem-solving approach to fuel innovation-led growth. Brüssel: European Commission. [https://ec.europa.eu/info/sites/info/files/mazzucato\\_report\\_2018.pdf](https://ec.europa.eu/info/sites/info/files/mazzucato_report_2018.pdf)
- Oberhofer, H., Schwarz, G., & Strassnig, M. (2019). Registerforschung: Verwaltungs- und Statistikdaten für die Wissenschaft. *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare* (72, 2). doi: 10.31263/voebm.v72i2.3154.
- OECD (2016). OECD Reviews of Innovation Policy: Sweden. Paris: OECD.
- OECD (2018). OECD Reviews of Innovation Policy: Austria. Paris: OECD.
- Ohler, F., Hutschenreiter, G., & Sturn, D. (1997). Evaluation of RTD-Programs: Best Practice considerations and the Austrian Experience. Newsletter Nr. 3 der Plattform FTEval, 2-12.
- Peneder, M. (1999). Creating a Coherent Design for Cluster Analysis and Relevant Policies. The Austrian "TIP" Experience. In OECD: OECD Proceedings. Boosting Innovation. The Cluster Approach, S. 339-359. Paris.
- Pichler, R., Stampfer, M., & Hofer, R. (2007). Forschung, Geld und Politik. Die staatliche Forschungsförderung in Österreich 1945–2005. Innsbruck/Wien/Bozen: Studienverlag.
- Rat für Forschungs- und Technologieentwicklung (2020). Bericht zur wissenschaftlichen und technologischen Leistungsfähigkeit Österreichs 2020. Wien.
- Schibany, A., Jörg, L., & Nones, B. (2005). Instrumente der Technologieförderung und ihr Mix. Technischer Bericht. Wien: InTeReg Research Report.
- Schot, J., & Steinmueller, W. (2018). Three frames for innovation policy: R&D, systems of innovation and transformative change. *Research Policy* (47), S. 1554-1567.
- Schwarz, G., König, K., Oberhofer, H., & Strassnig, M. (2020). Zugang zu Register- und Individualdaten für die wissenschaftliche Forschung in Österreich. *fteval JOURNAL for Research and Technology Policy Evaluation*, S. 11-15. doi: 10.22163/fteval.2020.464
- Streicher, J., Polt, W., & Unger, M. (2020). Eine Untersuchung der Marktsituation im Bereich der FTI-Evaluierung in Österreich. *fteval Journal for Research and Technology Policy Evaluation*, S. 72-81. doi: 10.22163/fteval.2020.472
- Turner, S., Liu, W., Klimek, P., & Cheong, S. (2020). The role of mainstreamness and interdisciplinarity for the relevance of scientific papers. *PLOS One* (April 2). doi: 10.1371/journal.pone.0230325
- Valovirta, V. et al. (2014). The Impact of Tekes Activities on Wellbeing and Environment. Helsinki: Tekes.
- Wu, L., Wang, D., & Evans, J. (2019). Large teams develop and small teams disrupt science and technology. *Nature* (556), S. 378-382.

## 6 Anmerkungen

- 1 Nicht Gegenstand dieser Erörterungen sind Evaluierungen von Organisationen oder Instrumenten, die im Wesentlichen auf informierten Vorort-Besuchen von Peers oder vergleichbaren Expert\*innen beruhen. Ebenso werden Evaluationsverfahren bei Publikationen hier nicht behandelt. In diesen Fällen treten die methodischen, finanziellen und weiteren Komplexitäten, die zu vorliegender Studie geführt haben, nicht oder nur in einem geringeren Ausmaß auf.
- 2 Ohler, Hutschenreiter, & Sturn (1997) beschreiben nicht nur Erfahrungen mit und Anforderungen von Programmevaluierungen, sie geben auch eine Reihe von Empfehlungen: nämlich (stärkere) Einführung von System- und Institutionenevaluierungen, Vergleichen und Lernen aus mehreren Evaluierungen auf nationaler und EU-Ebene, Verpflichtung zur Durchführung von Evaluierungen und klare, vorhersehbare Prozesse auf Auftraggeberseite, schließlich Verbesserungen auf der Methodenseite: Verbesserung der Daten- und Indikatorenlage; bessere empirische Methoden; Vorzug der Evaluierung von Effekten / Wirkungen gegenüber Prozesselementen. Das war 1997.
- 3 <https://cordis.europa.eu/programme/id/FP2-MONITOR/de>
- 4 Der Text basiert auf Streicher (2017), Warta und Philipp (2016), (Landsteiner, 2015), Reiner und Smoliner (2012) und Dinges & Schmidmayer (2010).
- 5 [https://www.parlament.gv.at/PAKT/VHG/XXVII/I/I\\_00239/index.shtml](https://www.parlament.gv.at/PAKT/VHG/XXVII/I/I_00239/index.shtml)
- 6 Archivierte Seite von <http://www.bmbwk.gv.at/forschung/fps/qual/uebersicht.xml>
- 7 Quelle für die Laufzeit: <https://www.derstandard.at/story/1289607885889/wissenschaftsbudget-institut-sora-von-sparmassnahmen-betroffen>
- 8 <http://www.tipconsortium.net/>
- 9 [https://www.nsf.gov/funding/pgm\\_summ.jsp?pims\\_id=501084](https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=501084)
- 10 <http://researchonresearch.org/>
- 11 Institut für Wissenschafts- und Technikforschung, Universität Wien (Ulrike Feld)
- 12 Institut für Technik- und Wissenschaftsforschung der Universität Klagenfurt (Daniel Barben) und Institut für Wissenschaftskommunikation und Hochschulforschung (Martina Merz)
- 13 <https://www.bmbf.de/foerderungen/bekanntmachung-2109.html>
- 14 <https://www.bmbf.de/foerderungen/bekanntmachung-1302.html>
- 15 <https://www.bmbf.de/foerderungen/bekanntmachung-2109.html>
- 16 <https://www.bmbf.de/de/wie-misst-man-fortschritt-4886.html>
- 17 <https://www.bmbf.de/de/forschung-und-innovation-transparenter-machen-9511.html>
- 18 <https://monitor.data4impact.eu/> weitere Projektbeteiligte: Public Policy and Management Institute (PPMI), ATHENA Research and Innovation Center in Information Communication & Knowledge Technologies (ARC), Qualia Technologies of Understanding SA, Consiglio Nazionale delle Ricerche (CNR), University of Boras
- 19 <http://www.tipconsortium.net/>.
- 20 Schweden war in Europa Vorreiter in der Postulierung der Missionsorientierung und der Grand Challenges. Die entsprechende Lund Deklaration stammt aus 2009, und ein paar Jahre später sind größere nationale Programme dazu aufgelegt worden.
- 21 Die Schwerpunkte und Themen werden hier unter <https://en.viablecities.se/fokusomraden-och-temanhttps://en.viablecities.se/fokusomraden-och-teman> ausführlicher beschrieben
- 22 <https://en.viablecities.se/nytt/european-viable-cities-day-2019>
- 23 <https://static1.squarespace.com/static/5dd54ca29c9179411df12b85/t/5f2a792b5b57746dd3ec6b80/1596619056823/Viable+Cities.pdf>
- 24 <http://researchonresearch.org>
- 25 <https://www.risis2.eu/risis-methods/>
- 26 <https://www.risis2.eu/about/>
- 27 Mobility Patterns and Career Paths of EU Researchers <https://www.more3.eu/>
- 28 Das IHS und das WIFO (als Vertreter der Plattform Registerforschung) führen aktuell eine Erhebung im Auftrag des BMBWF durch, mit dem Ziel, jene staatlichen Register zu erfassen, die für Forschungszwecke im Bereich der Sozial- und Lebenswissenschaften von Relevanz wären, sollte es gelingen, einen Zugang dazu zu ermöglichen. Diese Erhebung beruht auf einer Ersterfassung, die bereits 2018 durchgeführt wurde.

Dabei wurden zumindest 100 Register oder weitere staatliche Datenerfassungen genannt, die besonders forschungsrelevant sind.

- <sup>29</sup> Evidenzen dafür hat der WWTF aus seinen Aktivitäten in der Plattform Registerforschung bzw. im Rahmen von Interviews, die wir für Schwerpunkt- und Programmentwicklung geführt haben.
- <sup>30</sup> [https://www.tuwien.at/fileadmin/Assets/dienstleister/CSD/Digiprojekte/Abstract\\_RIS\\_synergy.pdf](https://www.tuwien.at/fileadmin/Assets/dienstleister/CSD/Digiprojekte/Abstract_RIS_synergy.pdf)
- <sup>31</sup> <https://www.elsevier.com/about/press-releases/corporate/elsevier-acquires-science-matrix-inc-provider-of-research-analytics-services-and-data>
- <sup>32</sup> <https://www.elsevier.com/icsr>
- <sup>33</sup> <https://www.wizdom.ai/>
- <sup>34</sup> <https://www.researchfish.net/>
- <sup>35</sup> <https://www.dimensions.ai/>; siehe auch (Hook, Porter, & Herzog, 2018)
- <sup>36</sup> <https://www.digital-science.com/>
- <sup>37</sup> Es sei hier erwähnt, dass die Projekte der beiden Ausschreibungen zur Indikatorik des BMBF leicht aufzufinden sind bzw. dass die Ergebnisse nicht gebündelt kommuniziert werden.
- <sup>38</sup> Viele Universitäten verfügen schon über durchaus avancierte interne Forschungsinformationssysteme, die umfangreiche Daten zu den Outputs der Forscher\*innen enthalten.
- <sup>39</sup> <https://bitbucket.org/lcharlin/tpms/src/master/>
- <sup>40</sup> Leider sind für dieses Programm die entsprechenden Informationen im Internet nicht mehr vorhanden, und es somit fast nur mehr in der Erinnerung einzelner Beteiligten existiert. Es gibt nur vereinzelte Quellen dazu, siehe z.B. [http://www.forschungsnetzwerk.at/downloadpub/Grundlagenpapier\\_gesellschaftsbezogener\\_forschung.pdf](http://www.forschungsnetzwerk.at/downloadpub/Grundlagenpapier_gesellschaftsbezogener_forschung.pdf)
- <sup>41</sup> Im universitären Bereich werden diese über interne Programme (z.B. WU Wien) oder über eingeworbene Drittmittel im Rahmen von Projekten finanziert. Für die USA bis zum Doktorat gibt es im naturwissenschaftlich-technischen Bereich Fellowships bei der Austrian Marshall Plan Foundation.
- <sup>42</sup> <https://www.ffg.at/stiftungsprofessur>