

rütter soceco

sozioökonomische forschung + beratung

**senarclens
leu+partner**

strategische kommunikation



Quantum
Business Intelligence Services AG

rütter soceco ag
Weingartenstrasse 5
CH-8803 Rüschlikon
Tel. (+41) 044 724 27 70
Fax (+41) 044 724 22 78
info@ruetter-soceco.ch
www.ruetter-soceco.ch
Partner von:
[senarclens, leu + partner ag](#)

Machbarkeitsstudie zur Einrichtung einer österreichweiten Forschungsförderungsdatenbank

Bericht für den Rat für Forschung und Technologieentwicklung

Zürich, 10.11.17

Auftraggeber

Rat für Forschung und Technologieentwicklung
Pestalozzigasse 4
1010 Wien

Projektleitung

Oliver Hoff
Rütter Soceco AG
Sozioökonomische Forschung + Beratung
Weingartenstrasse 5
8803 Rüschlikon
Tel.: +41 (0) 44 724 27 70
Email: oliver.hoff@ruetter-soceco.ch

Autoren

Oliver Hoff, Rütter Soceco
Corina Rieser, Rütter Soceco
Dr. Andrea Leu, Senarclens Leu & Partner
Myriam Hofmann, Senarclens Leu & Partner
Ole Vossnack, Quantum Analytics
Nadim Schumann, Quantum Analytics

Dieser Bericht ist ausschliesslich zur Information des Auftraggebers bestimmt.
Jede Weitergabe an Dritte bedarf der vorgängigen schriftlichen Zustimmung durch
Rütter Soceco AG.

Inhaltsverzeichnis

1. Ausgangslage und Fragestellung	4
2. Ziele und Methoden	5
3. Übersicht: Daten der Forschungsförderung in Österreich	6
3.1 Schwerpunktlegung	6
3.2 Die Struktur der Forschungsförderung in Österreich	7
3.2.2 Bund und Länder	12
3.2.3 Intermediäre	13
3.2.4 Empfänger von Fördermitteln	15
4. Anwendungsmöglichkeiten und Ansprüche	15
4.1 Nutzen und Anwendungsfälle einer F(F)DB	15
4.2 Kerndatensatz	17
4.2.1 Datenschutz und rechtliche Fragen	20
4.2.2 Datenstandardisierung	21
5. Varianten für die Implementation einer Datenbank	23
5.1 Variante 1: Meta-Projektfinder ohne weitere Datenspeicherung	24
5.2 Variante 2: Ausbau Transparenzdatenbank	25
5.3 Variante 3: Aufbau einer neuen Graphdatenbank	27
5.4 Variante 4: EU-weite Vernetzung	28
5.5 Die Varianten im Vergleich	31
6. Internationale Case Studies	37
6.1 Case Study: Schweiz	37
6.1.1 Institutionen der Forschungsförderung	37
6.1.2 Forschungsförderungs-Datenbanken	38
6.1.3 Daten	39
6.2 Case Study: Deutschland	42
6.2.1 Institutionen der Forschungsförderung	42
6.2.2 Forschungsförderungs-Datenbanken	42
6.2.3 Daten	43
6.3 Case Study: Neuseeland	45
6.3.1 Institutionen der Forschungsförderung	45
6.3.2 Forschungsförderungs-Datenbanken	45
6.3.3 Daten	46
7. Fazit	48
8. Anhang	49
8.1 Liste Der Gesprächspartner (national)	49
8.2 Liste Der Gesprächspartner (international)	50

1. Ausgangslage und Fragestellung

Die gesamten Mittel für Forschung und Entwicklung (F&E) in Österreich haben sich im Jahr 2015 auf insgesamt rund 10.1 Mrd. EUR belaufen. Davon gehen rund 3.8 Mrd. EUR auf Beiträge des öffentlichen Sektors zurück (RH, 2016). Bereits 2011 hat sich die österreichische Bundesregierung zum Ziel gesetzt, ihre Forschungsquote¹ bis zum Jahr 2020 auf 3.76% zu steigern, um damit zur europäischen Spitze aufzuschliessen. 2015 hat die Bundesregierung eine Quote von 3.01% erreicht. Es besteht jedoch seit mehreren Jahren das Problem, dass es keine umfassende und zentrale Übersicht über öffentliche Forschungsförderungsmittel, -programme und -projekte gibt. Vor dem Hintergrund der angestrebten Entwicklung im Bereich F&E wird dieses Problem daher zunehmend akuter.

Der Rechnungshof kommt in seinem Bericht 2016 zur Erkenntnis, dass es überlappende Mehrfachstrukturen im Bereich der öffentlichen Forschungsförderung gibt. So gibt es neben den 216 Organisationseinheiten auf Bundes- und Länderebene noch 24 Intermediäre, die mit F&E-Angelegenheiten betraut sind sowie 136 Forschungsprogramme mit unterschiedlichen Forschungszielen (RH, 2016 S. 269). Zahlreiche Datenbanken auf unterschiedlichen Ebenen (Bund, Länder, Intermediäre), die zu unterschiedlichem Zweck betrieben werden, erschweren den Überblick zusätzlich (RH, 2016, S. 257).

Um diese Situation zu adressieren, hat der Rechnungshof 2005 und 2016 erneut die öffentliche Forschungsfinanzierung in Österreich untersucht, inventarisiert und einen Überblick der Forschungsfinanzierung erstellt (vgl. Abbildung 1).

Es gibt verschiedene Formen der Forschungsförderung, die wiederum von einer ganzen Reihe von Akteuren verwaltet werden. Eine umfangreiche und vertiefte Darstellung des Systems findet sich im Rechnungshofbericht 2016. In diesem Bericht werden die relevanten Erkenntnisse des Rechnungshofberichts in einer Form zusammengefasst und dargestellt, die das Thema Forschungsförderungsinformationen greifbar macht.

Aus einer früheren Version dieses Berichts des Rechnungshofs aus dem Jahr 2005 geht hervor: *„In Österreich war eine flächendeckende Erhebung von Forschungsförderungsdaten mangels einer den Bund und die Länder umfassenden Forschungsförderungsdatenbank nicht möglich. Eine österreichweite Gesamtschau nach geförderten Projekten und Förderungsnehmern fehlte. (...)“*

Die vielfältigen Bestimmungen zum Thema Forschungsförderung führten zu zahlreichen Parallelerhebungen von Forschungsdaten und je nach Erhebung zu unterschiedlichen Förderungsbegriffen. Ebenso erfolgte die Erfassung der Forschungsprojekte sowie der jeweiligen Förderungsnehmer zumeist dezentral bei den einzelnen Förderungsstellen und nach unterschiedlichen Kriterien. (...)“

Das BMVIT und das BMBWK verfügten über keine durchgängigen, IT-unterstützten Forschungsförderungsdatenbanken. Förderungsrelevante Informationen über Förderungsnehmer bezogen die Ressorts aus einer Vielzahl von Datenquellen. Die damit befassten Organisationseinheiten erstellten teils selbst eigene Individuallösungen zur Datenverarbeitung mit unterschiedlicher Qualität. Unerwünschte Mehrfachförderungen ein und desselben Förderungsnehmers oder Forschungsprojekts konnten von den Ressorts nicht mit Sicherheit vermieden werden“ (S. 21f).

Rund zehn Jahre später zieht der Rechnungshof die folgenden Schlüsse hinsichtlich des Themas Forschungsförderungsdatenbank (FFDB):

¹ Der Anteil der Forschungsausgaben am Bruttoinlandsprodukt

² Für eine Übersicht siehe z.B.: https://en.wikipedia.org/wiki/List_of_academic_databases_and_search_engines

³ Die Diskrepanz zwischen Bezeichnung und Domain ergibt sich durch die Überschneidung mit der Transpa-

„Bei den Forschungsförderungsdatenbanken bestanden – ungeachtet der Transparenzdatenbank des Bundes und der Bundesforschungsdatenbank des BMWF – beim Bund vier weitere Datenbanken sowie in den Ländern 14 Datenbanken, welche primär auf die Erfassung und Abwicklung von Forschungsförderungsmaßnahmen und/ oder auf Informationsbedürfnisse zugeschnitten waren. Diese Datenbanken erfüllten zwar die operativen Bedürfnisse, boten jedoch nicht die Möglichkeit zur länder- bzw. bund-länderübergreifenden Informationsbeschaffung oder zur Abbildung der gebietskörperschaftsübergreifenden Zahlungsströme in der Forschungsfinanzierung“ (2016, S. 30).

Der Rechnungshof (2016) empfiehlt daher:

„Mit einer einheitlichen Datenbank sollte sichergestellt werden, dass die Forschungsförderungsleistungen österreichweit umfassend und vollständig erfasst werden. (...)

„Auf der Grundlage von Kosten-Nutzen-Analysen sollte geprüft werden, ob eine Integration der bestehenden länder- oder bundes-spezifischen Datenbanken in eine umfassende österreichweite Forschungsförderungsdatenbank möglich wäre. (...)

Für den Fall, dass eine vollständige Integration der länder- oder bundesspezifischen Datenbanken in eine umfassende österreichweite Forschungsförderungsdatenbank nicht möglich ist, wäre durch Schnittstellen die Möglichkeit der Datenübertragung sicherzustellen. (...)

Im Falle der Realisierung einer eigenständigen, umfassenden Forschungsförderungsdatenbank oder einer vergleichbaren IT- Lösung wäre die Bundesforschungsdatenbank in diese zu integrieren und nicht mehr als gesonderte Datenbank zu führen“ (S. 147).

2. Ziele und Methoden

Ziele

Vor dem Hintergrund der aktuellen dargestellten Situation wurde die hier vorliegende *Machbarkeitsstudie* zur Einrichtung einer Forschungsförderungsdatenbank erstellt. Diese hat die folgenden Ziele:

- Erstellen einer Übersicht über die wichtigsten FFDBs in Österreich sowie deren Struktur und Inhalt.
- Untersuchung ausgewählter internationaler Case Studies
- Aufzeigen der Verwendungsmöglichkeiten einer FFDB vor dem Hintergrund politischer, inhaltlicher, öffentlicher und institutioneller Interessen.
- Aufzeigen von technischen Lösungen für die Integration der Daten aus verschiedenen Quellen bzw. von Schnittstellen zwischen verschiedenen Datenbanken.
- Skizzierung eines Kerndatensatzes für Einträge in die Datenbank und der dafür notwendigen Standardisierungen existierender Daten.
- Erstellen einer Übersicht über die datenschutzrechtlichen Fragestellungen, die mit dem Betrieb einer derartigen Datenbank verknüpft sind.
- Einschätzung über die zu erwartenden Kosten für die technische Umsetzung und für den laufenden Betrieb einer solchen FFDB.

Explizit *nicht* Teil dieser Untersuchungen sind weitere durch den Rechnungshof aufgeworfene Fragen wie z.B. die Möglichkeit ungewollter Mehrfachförderungen, redundante oder überlappende Förderungsstrukturen, inhaltliche und organisatorische Forschungs-koordination zwischen Bund und Ländern sowie komplexe Förderungsstrukturen.

Verwendete Methoden

Zur Erreichung der oben genannten Ziele wurden im Zeitraum von Mai bis August 2017 eine Reihe von semi-strukturierten Interviews mit Vertretern und Vertreterinnen von relevanten Forschungsförderungsinstitutionen in Österreich geführt. Sie bilden die Grundlage dieses Berichts (vgl. Anhang). Ergänzt wurden die aus diesen Interviews gewonnen Erkenntnisse mit einer Literaturanalyse, die sowohl Forschungsliteratur zum Thema Forschungsdatenorganisation als auch Jahresberichte, Strategiepapiere, Konzeptstudien, etc. beinhaltete. Aufbauend auf den Interviews und der Literaturanalyse zur Situation in Österreich wurden anschliessend Interviews mit Repräsentanten relevanter Forschungsförderungsinstitutionen im Ausland geführt. Ziel war es, zu untersuchen, wie andere Länder mit vergleichbaren Herausforderungen umgehen. Es wurden Gespräche mit Vertretern aus Deutschland, der Schweiz und Neuseeland geführt.

3. Übersicht: Daten der Forschungsförderung in Österreich

Aufbauend auf den Untersuchungen des Rechnungshofes (vgl. RH, 2005, 2016) wird in Österreich im Rahmen der Weiterentwicklung der Strategie für Forschung, Technologie und Innovation“ (FTI-Strategie) nach Wegen gesucht, durch bessere Datenhandhabung die Koordination und Evaluation der öffentlichen Forschungsförderung zu verbessern. Eine Möglichkeit dafür besteht in der Einrichtung einer gebietskörperschaftsübergreifenden FFDB. Diese Untersuchung beschäftigt sich mit den Fragen:

- was unter dem Begriff „Forschungsförderungsdatenbank“ zu verstehen ist,
- welche Funktionen eine derartige Datenbank zu erfüllen vermag und welche nicht,
- welche Inhalte notwendig sind,
- welche Szenarien es gibt, eine derartige Datenbank im aktuellen FTI-System Österreichs zu verankern und
- mit welchen Aufwänden diese Szenarien verknüpft sind.

Um diese Fragen zu beantworten, wird im Bericht zunächst die aktuelle Situation in Österreich untersucht und dargestellt (Kapitel 3). Anschliessend werden im Kapitel 4 Anwendungsfälle und Nutzern einer möglichen FFDB skizziert. Auch werden in diesem Kapitel datenschutzrechtliche Fragen diskutiert sowie ein Kerndatensatz für Forschungsprojekte vorgestellt. Kapitel 5 geht dann auf vier Varianten ein, wie sich eine FFDB in Österreich (bzw. auf EU-Ebene) realisieren lassen würde. Anschliessend werden in Kapitel 6 die Ergebnisse der internationalen Case Studies vorgestellt und im Kapitel 7 ein Fazit gezogen.

3.1 Schwerpunktlegung

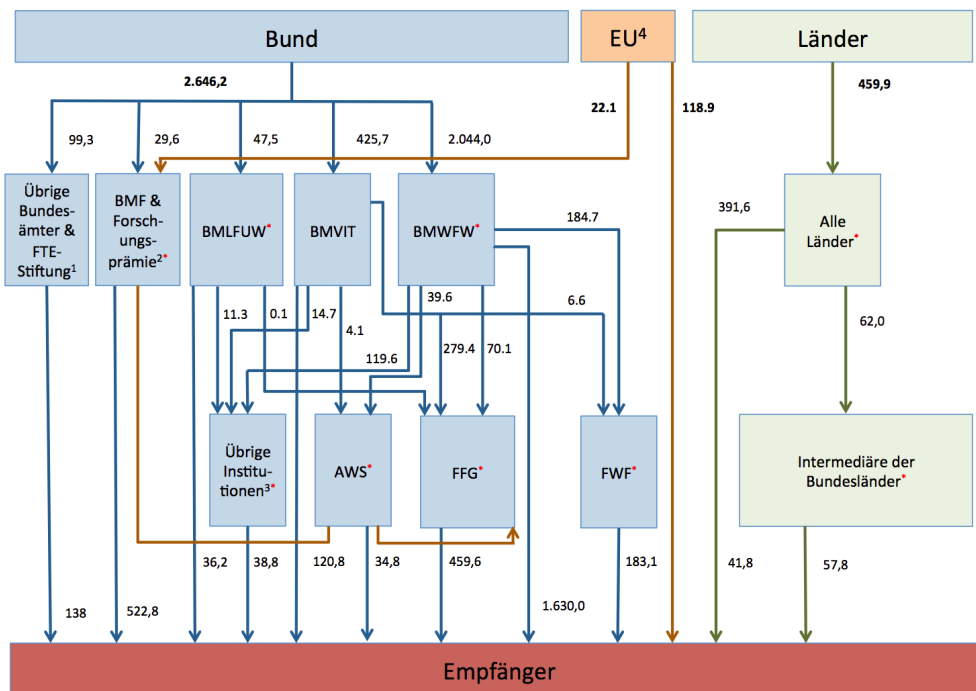
Wie der Rechnungshof in seinen Untersuchungen des Forschungsförderungssystems Österreichs festgestellt hat, sind die Strukturen und Finanzflüsse komplex. Um das Thema fassbar zu machen, wurden daher in Abstimmung mit dem Auftraggeber dieser Untersuchung – dem Rat für Forschung und Technologieentwicklung – bestimmte Schwerpunkte gesetzt, die in diesem Bericht vorrangig untersucht werden. Auf organisatorischer Ebene liegt der Schwerpunkt dieses Berichts daher auf der *Forschungsför-*

derung auf Bundesebene. Die liegt darin begründet, dass finanziell gesehen der weit- aus grösste Teil der gesamten öffentlichen Forschungsförderung über Bundesinstitutio- nen abgewickelt wird (vgl. Abbildung 1).

Die öffentliche Forschungsförderung der Bundesländer wird im Bericht ebenfalls be- handelt. Die Bundesländer werden jedoch trotz regionaler Unterschiede als homogene Gruppe adressiert. Der Grund dieser Gruppierung ist zum einen eine Reihe von Ge- meinsamkeiten (v.a. hinsichtlich relevanter Faktoren für eine FFDB) trotz regionaler Unterschiede eine Reihe von Gemeinsamkeiten aufweisen und zum anderen die mo- dulare Konstruktion der Datenbankszenarien (vgl. Kapitel 5). Konkret bedeutet dies, dass die Skalierung einer möglichen Datenbanklösung mit zusätzlichen Institutionen mit relativ wenig Aufwand möglich ist, auch wenn in einem ersten Schritt evtl. nur wenige zentrale Institutionen einbezogen werden sollten. Hinzu kommt, dass die verschiedenen Szenarien, die für die Einrichtung einer FFDB vorgeschlagen werden, *modular* konstru- iert sind, d.h. die Skalierbarkeit flexibel aufgebaut ist. Weitere oder andersartige Daten- quellen können ohne grossen Aufwand angehängt werden.

3.2 Die Struktur der Forschungsförderung in Österreich

Um aufzuzeigen, wie sich die Forschungsförderung in Österreich strukturiert, ist ein Verständnis der Akteure und Geldflüsse nötig. Eine umfangreiche und vertiefte Darstel- lung der Flüsse findet sich im Rechnungshofbericht 08/2016. In Abbildung 1 sind diese Informationen so zusammengefasst und dargestellt, dass sie das Thema Forschungs- förderungsinformationen greifbar machen (vgl. zur Erläuterung auch das Abkürzungs- verzeichnis).

Abbildung 1: Schematische Darstellung der Forschungsförderung in Österreich (2014)

Zudem gilt es unterschiedliche *Arten der Forschungsförderung* zu unterscheiden, um ein Verständnis für die Anwendungsfälle einer FFDB zu ermöglichen. Grundsätzlich lässt sich die gesamte öffentliche Forschungsförderung in *drei Komponenten* unterteilen

1. Grundlagenfinanzierung des Hochschulbereichs (50%)
2. Ausserhochschulische institutionelle Globalförderungen, Mitgliedschaften (z.B. Mitgliedsbeiträge an die Europäische Weltraumorganisation), ministeriale Ad-hoc Dienstleistungsaufträge mit Forschungscharakter sowie Personal und Verwaltungsaufwand (25%)
3. Programmförderungen
 - Bundesprogramme (23%)
 - Länderprogramme (2%)

In Klammern ist dabei jeweils angegeben, wie hoch der Anteil der jeweiligen Komponente an der *gesamten* öffentlichen Forschungsförderung in Österreich ist. Dabei ist zu beachten, dass es sich dabei um eine grobe Schätzung handelt.

Der Anspruch der vorliegenden Untersuchung ist es, die Machbarkeit einer FFDB zu untersuchen. Zum Verständnis der folgenden Ausführungen ist es daher notwendig, zunächst zu definieren, was mit einer FFDB gemeint ist und was sie beinhalten kann. Eine FFDB in diesem Bericht ist eine *IT-Lösung*, die den *Zugang zu Informationen der öffentlichen Forschungsförderung* ermöglicht. Forschungsförderungsinformationen haben dabei eine *weite und eine enge Definition*. Die *weite Definition* enthält *alle* Informationen, die von der Finanzierungs- über die Durchführungs- bis zur Empfängerebene in allen drei Komponenten anfallen. Die engere Definition grenzt diese Informationsmenge stark ein. In der *engen Definition* sind ausschliesslich Informationen zu jenen *Forschungsprojekten* gemeint, die im Rahmen von *Forschungsprogrammen* in Form von Drittmitteln teilweise oder gänzlich durch Gelder der öffentlichen Hand gefördert werden. Diese Untersuchung baut in Abstimmung mit dem Rat für Forschung und Technologieentwicklung auf diese enge Definition von Forschungsförderungsinformationen auf.

Eine FFDB bezeichnet im Folgenden *eine IT-Lösung, die den Zugang zu Informationen von Forschungsprojekten ermöglicht, die teilweise oder gänzlich mit Geldern der öffentlichen Hand finanziert wurden*. Dies bedeutet, dass vor allem Informationen der dritten Gruppe sowie vereinzelt Informationen der zweiten Gruppe (Ad-hoc Forschungsaufträge der Bundesämter) für diese Untersuchung relevant sind. Da weiterhin rund 90% der Gelder der *Bundesprogrammforchung* über die drei grossen Förderungsintermediäre Fonds zur Förderung der wissenschaftlichen Forschung (FWF), Österreichische Forschungsförderungsgesellschaft mbH (FFG) und Austria Wirtschaftsservice GmbH (AWS) abgewickelt werden, bildet die Betrachtung dieser Institutionen bzw. die Betrachtung ihrer Datenstruktur und -verwendung den Kern dieses Berichts.

Forschungsdaten und Forschungsförderungsdaten

Sowohl der Rechnungshof als auch die Zielsetzung dieser Untersuchung beschränken sich auf die Untersuchung der Organisations- und Darstellungsform von *Forschungsförderungsinformationen*. Generell gesehen, haben sie mit *Forschungsinformationen* viel gemeinsam (vgl. Kapitel 4.2). Es bestehen jedoch auch spezifische Unterschiede. *Zum einen liegt der inhaltliche Fokus* bei Forschungsförderungsinformationen klar auf dem Förderungsaspekt. Das bedeutet, dass die zentralen Informationselemente einer Forschungsförderungsinformation immer in Bezug zur Forschungsförderung stehen (z.B. Finanzielles, Zeiträume, Akteure). Das schliesst nicht aus, dass auch andere Informati-

onen Teil einer Forschungsförderungsinformation sein können. Jedoch werden diese anderen Forschungsinformationen nur durch die Verknüpfung mit den *Förderungsinformationen* zu Forschungsförderungsinformationen.

Die *zweite Unterscheidung* liegt in der *institutionellen Zuordnung von Forschungsförderungsinformationen im Vergleich zu Forschungsinformationen*. Forschungsförderungsinformationen werden traditionell vor allem bei jenen Institutionen erhoben, gespeichert und nutzbar gemacht, die Förderungen vergeben bzw. mit der operativen Abwicklung von Förderung betraut sind. In Österreich sind dies primär die Forschungsintermediäre (vgl. Hinweis 2 in Abbildung 1), die mit der operativen Abwicklung von Forschungsprojekten beauftragt sind. Auch andere Institutionen (z.B. Universitäten in Publikationsdatenbanken oder Forschungsrepositorien) speichern in ihren internen Datenbanken häufig Forschungsförderungsinformationen als ergänzende Informationen. Sie stehen in diesen Datenbanken jedoch aufgrund des Zwecks der Datenbank und der institutionellen Verankerung nicht im Zentrum. Zudem werden in Datenbanken von solchen nicht direkt auf Förderung ausgerichteten Institutionen verschiedene Datensätze gesammelt, die eine Abgrenzung zwischen Forschungsinformationen und Forschungsförderungsinformationen notwendig machen.

An einem Beispiel lässt sich diese inhaltliche und institutionelle Unterscheidung am besten verstehen. Traditionelle Systeme mit Forschungsinformationen (Forschungsinformationssysteme) sind z.B. Publikations- oder Projektdatenbanken, die von den meisten Hochschulen und Forschungsinstitutionen betrieben werden. Diese Datenbanksysteme erfassen und speichern Daten, zu *Publikationen (oder Projekten)*, die von Universitätsangehörigen verfasst oder durchgeführt werden. Dabei ist es wichtig zu berücksichtigen, dass diese Datenbanken einerseits Publikationen enthalten, die im Rahmen von Forschungsprojekten entstanden, die mit öffentlichen Drittmitteln gefördert wurden, die über die Grundlagenfinanzierung der Hochschulen hinausgehen (Forschungsförderungsinformationen), andererseits jedoch auch Publikationen, die in Projekten ohne derartige Drittmittelförderung entstanden sind (Forschungsinformationen). Gleichzeitig ist es jedoch auch möglich, dass ein an einer Hochschule angestellter Forscher ein öffentlich drittmittelgeförderter Forschungsprojekt abwickelt, dass keine damit verknüpfte Publikation aufweist. Der Datensatz dieses Forschungsprojekts würde wiederum nur in den Datenbanken der fördernden Institution auftauchen, nicht jedoch in der Publikationsdatenbank der Universität, da deren Fokus i.d.R. auf Publikationen liegt und nicht auf Forschungsprojekten im Allgemeinen. Es zeigt sich also, dass es zwischen den aktuellen Speicherorten von Forschungsförderungsinformationen und den Speicherorten von Forschungsinformationen zum einen Doppelspurigkeiten gibt (Datensätze zu identischen Projekten an unterschiedlichen Orten). Gleichzeitig gibt es jedoch auch Datenlücken bei den Betreibern von Forschungsinformationssystemen, die teilweise einen manuellen Bezug von Daten von Forschungsintermediären erfordern.

Die *dritte Unterscheidung* zwischen Forschungsförderungsinformationen und Forschungsinformationen bzw. zwischen den jeweils dafür erstellten Datenbanksystemen liegt in ihrem *Verwendungszweck*. Forschungsförderungsinformationen werden von Forschungsintermediären primär für die operative Abwicklung von Forschungsprogrammen (von der Antragsstellung bis zur Archivierung) verwendet. Zudem werden Forschungsförderungsinformationen auf politischer Ebene für die Planung, Überprüfung und Evaluation der Wissenschaftspolitik des Bundes genutzt. Hinzu kommen Verwendungsfälle durch die Rechenschaftspflicht der Forschungsintermediäre (Transparenzanspruch), Berichtserstellung sowie ad-hoc Anfragen von Hochschulen, Forschenden, Journalisten, Politikern und anderen interessierten Stellen. Forschungsinformationen bzw. Forschungsinformationssysteme werden von Forschungsinstitutionen hingegen vor allem für die Sammlung, Darstellung und Weiterentwicklung ihrer eigenen For-

schungsleistung betrieben (z.B. für die Berichtspflichten im Rahmen der Wissensbilanzen). Zudem werden sie von Forschenden für die Sammlung und Darstellung ihrer persönlichen Forschungsleistung genutzt, zur Recherche möglicher Projektpartner und Projektideen, sowie von Firmen und Medien für die Suche nach Experten und Forschungspartnern. Institutionelle FIS sind gemäss ihrer Inhalte und Spezifikationen eng verknüpft mit internationalen, teilweise fachspezifischen Publikationsdatenbanken wie z.B. Scopus (ELSEVIER), Web of Science (Thomson & Reuters), Google Scholar (Alphabet) und vielen weiteren².

Vor dem Hintergrund dieser Unterschiede zwischen Forschungsförderungsinformationen und Forschungsinformationen wird klar, dass es für die Diskussion einer strukturübergreifenden, nationalen Lösung von grosser Bedeutung ist, zu untersuchen, ob ausschliesslich Interesse an der Sammlung von *Forschungsförderungsinformationen* besteht oder ob die Sammlung und Nutzbarmachung der *gesamten* im System vorhandenen *Forschungsinformationen* nicht zielführender wäre. Bei der Variante einer FFDB wird eine Übersicht über Forschungsprojekte mit öffentlicher Programmförderung gewonnen. Die Variante eines *nationalen Forschungsinformationssystems (NFIS)* hingegen würde eine Übersicht über die *gesamte Forschungslandschaft* an österreichischen Forschungsinstitutionen ermöglichen, inklusive der durch öffentliche Programmförderung geförderten Forschungsprojekte. Sowohl die Literaturanalyse, als auch die internationalen Case Studies haben ergeben (vgl. Kapitel 6), dass sowohl die Entwicklungen in der internationalen Forschungsgemeinschaft, als auch in den nationalen FTI-Systemen in beide Richtungen gehen können. Sowohl NFIS Ansätze werden verfolgt (Norwegen, Niederlande) als auch FFDB Ansätze (Grossbritannien, Neuseeland). Hinzu kommen Ansätze, die *ausschliesslich* von Hochschulen getragen werden und Forschungsförderungsinformationen daher nur indirekt und nicht institutionell abbilden (vgl. Dänemark, Schweiz). Tabelle 1 zeigt verschiedene Beispiele:

Tabelle 1: Übersicht internationale NFIS Ansätze

Indikator	Norwegen	Grossbritannien	Niederlande	Neuseeland	Dänemark
Name	CRISTin	Gateway to Research	NARCIS	NRIS	Danske Forskningsdatabaser
URL	cristin.no	gtr.rcuk.ac.uk	narcis.nl	Im Aufbau	forskningsdatabaser.dk
Personenbezogene Informationen	Ja	Ja	Ja	Ja	Ja
Daten von Förderungsinstitutionen	Ja	Ja	Ja	Ja	Nein
Daten von Forschungsinstitutionen	Ja	Nein	Ja	Teilw.	Ja

Quelle: Rütter Soceco, eigene Darstellung

Die in dieser Untersuchung vorgenommene Analyse konzentriert sich gemäss der Zielsetzung auf die erste Variante (Forschungsförderungsdatenbank). Die zweite Variante (NFIS) ist *nicht* im Fokus dieser Untersuchung, da erhebliche Unterschiede in der Organisationsstruktur, im Standardisierungsbedarf, den Anwendungsfällen, im Nutzen und bei den Aufwänden zwischen den Varianten bestehen. Dennoch wird – soweit möglich – in diesem Bericht an verschiedenen Stellen darauf hingewiesen, dass die verschie-

² Für eine Übersicht siehe z.B.: https://en.wikipedia.org/wiki/List_of_academic_databases_and_search_engines

denen Überlegungen und Analyseschritte grundsätzlich für beide Varianten gelten. Zu diesem Zweck wird die Abkürzung F(F)DB verwendet, um zu verdeutlichen, dass der Gedankengang auch für eine Forschungsdatenbank anstelle einer Forschungsförderungsdatenbank wichtig ist.

In den folgenden Kapiteln wird als Grundlage für die Entwicklung der Datenbankvarianten kurz dargestellt, bei welchen Institutionen in Österreich aktuell welche Daten zu welchem Zweck gespeichert werden.

3.2.2 Bund und Länder

Bundesministerium für Wissenschaft, Forschung und Wirtschaft (BMWFW)

- Betreibt seit den 1970er Jahren die Bundesforschungsdatenbank (B_f.dat)
- Die Datenbank erfasst institutionelle Mitgliedschaften und ad-hoc Dienstleistungs- und Forschungsaufträge der österreichischen Bundesministerien.
- Etwa 85% der Fördermittel entfallen auf Globalförderungen an Forschungsinstitutionen oder Förderagenturen
- Beinhaltete 2016 rund 633 laufende F&E-Projekte des Bundes
- Nicht alle in der B_f.dat enthaltenen Projekte haben einen Forschungscharakter, dies ist bei der Integration der Daten zu berücksichtigen.
- Meldepflicht der Bundesministerien ist im Forschungsorganisationsgesetz (FOG) geregelt.
- Trotz Meldepflicht gehen die Betreiber davon aus, dass keine vollständige Abdeckung aller F&E-Projekte aller Ministerien/Ressorts vorliegt. Grundsätzlich besteht jedoch hohe Abdeckung.

Bundesministerium für Finanzen (BMF)

- Betreibt transparenzportal.gv.at (ugs. „Transparenzdatenbank³“)
- Diese Datenbank wurde geschaffen, um Transparenz in der öffentlichen Förderung von natürlichen und juristischen Personen herzustellen.
- Natürliche und juristische Personen können in dieser Datenbank ihre erhaltenen Leistungen und die für Sie möglichen Förderungen einsehen und teilweise beantragen.
- Deckt gesamtes Spektrum des öffentlichen Leistungs-/Förderangebots ab, von Wohnbeihilfen bis Kulturförderung.
- Hat eine eigene rechtliche Grundlage. Transparenzdatenbankgesetz (TDBG 2012).
- Hat rigides Authentifizierungssystem und Sicherheitsprotokolle, da sensible und personenbezogene Daten enthalten und abgerufen werden können. Individuelle Zugänge für Bürger, Firmen und andere Institutionen.
- Gemäss dem Stand 2017 sind alle Leistungsangebote auf Bundes- und Länderebene erfasst. Leistungsbezüge sind mit wenigen Ausnahmen nur auf Bundesebene erfasst.

³ Die Diskrepanz zwischen Bezeichnung und Domain ergibt sich durch die Überschneidung mit der Transparenzdatenbank (transparenzdatenbank.at) des BMLFUW, die Agrarsubventionen der EU sammelt und zugänglich macht. In diesem Bericht ist mit Transparenzdatenbank (TDB) grundsätzlich die Lösung des BMF gemeint.

- Im Bereich Forschungsförderung sind mit entsprechender Berechtigung und auf Leistungsbezugsebene ausschliesslich Informationen zu Empfänger, Zahlstelle, Betrag und Betreff abrufbar.
- Es sind nicht alle Zahlungen im Bereich der Forschungsförderung in Österreich erfasst. Es ist bspw. offen, ob Zahlungen bzw. Projekte, die in der B_f.dat erfasst sind, aufgrund ihres Ad-hoc Charakters ebenfalls in der Transparenzdatenbank erfasst sind. Das Gleiche gilt für Zahlungen der Bundesländer und deren Intermediäre.
- Der Zweck und der Funktionsumfang der Transparenzdatenbank sind aktuell relativ weit von den Ansprüchen einer FFDB entfernt und würden erheblichen Anpassungen bedürfen.
- Es ist fraglich, ob die umfassenden Anpassungen an der Transparenzdatenbank ausschliesslich für den Bereich Forschungsförderung möglich sind, ohne die weitreichenden anderen Ansprüche an die Transparenzdatenbank zu beeinflussen. Es ist jedoch denkbar, die Transparenzdatenbank generell umzustrukturieren, um den Bedürfnissen der anderen öffentlichen Stellen entgegenzukommen.
- Sollte eine mögliche Revision zum Tragen kommen, ist zu prüfen, ob die Leistungsvereinbarungen zwischen den Ministerien anzupassen sind bzw. ob das BMF die richtige Institution ist, eine um die Funktionalitäten einer FFDB ergänzte Transparenzdatenbank zu betreiben.

Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft (BMLFUW)

- Betreibt seit 2005 die Datenbank für Forschung zur nachhaltigen Entwicklung (DaFNE)
- Basiert auf dem Common European Research Information Format (CERIF) der International Organisation for Research Information
- Wird für BMLFUW interne Anträge und für Verarbeitung von externen Projektanträgen sowie zur Veröffentlichung von Projektinformationen und Forschungsergebnissen verwendet.
- Wird für Projekte des BMLFUW verwendet (v.a. Land- und Forstwirtschaft, nachhaltige Entwicklung)

Es gibt noch andere Bundesministerien, die eigene Datenbanken für Forschungsprojekte betreiben (z.B. Bundesministerium für Inneres, Bundesministerium für Landesverteidigung und Sport). Da diese jedoch einen sehr geringen Umfang haben und thematisch sehr spezifisch sind, werden Sie an dieser Stelle nicht weiter ausgeführt.

Die österreichischen Bundesländer bzw. deren Intermediäre betreiben eine ganze Reihe von Datenbanken (13, gemäss Prüfung des Rechnungshofes), die für die Erfassung und Abwicklung von Forschungsprojekten eingesetzt werden. Eine detaillierte Übersicht dieser Datenbanken findet sich im angesprochenen Bericht des Rechnungshofes (2016, S. 257).

3.2.3 Intermediäre

FWF – Der Wissenschaftsfonds

- Ist zuständig für die operative Abwicklung von Grundlagenforschungsprogrammen. Vergibt Forschungsgelder grösstenteils an Universitäten.

- Betreibt eine interne SQL-Datenbank mit ausführlichen Informationen zu Forschungsprogrammen, -projekten, Personen, etc.
- Diese Datenbank dient als zentrales Werkzeug für die Erfüllung des Leistungsauftrags des FWF und wird u.a. für die Abwicklung von Forschungsprogrammen und -projekten sowie für die Berichtslegung und Rechenschaftspflichten nach aussen genutzt.
- Zentrale Projektinformationen werden für genehmigte, laufende und abgeschlossene Projekte über eine Webschnittstelle öffentlich verfügbar gemacht (auch rückwirkend).
- Veröffentlichung der Daten basiert auf freiwilliger Zustimmung der Antragsteller.
- Berücksichtigt bei der Erfassung von Publikationen und Metadaten internationale Datenstandards.

FFG – Österreichische Forschungsförderungsgesellschaft

- Hat als Leistungsauftrag die Förderung von Forschungs- und Entwicklungsvorhaben natürlicher und juristischer Personen sowie die Förderung der Kooperation zwischen Wissenschaft und Wirtschaft zum Ziel.
- Betreibt ebenfalls eine interne SQL-Datenbank mit ausführlichen Informationen zu Forschungsprogrammen, -projekten, Personen, etc.
- Diese Datenbank dient als zentrales Werkzeug für die Erfüllung des Leistungsauftrags der FFG und wird u.a. für die Abwicklung von Forschungsprogrammen und -projekten genutzt sowie für die Berichtslegung und Rechenschaftspflichten nach aussen
- Datenauszüge der Datenbank werden zudem für interne und externe Evaluationen ausgewertet sowie für das regelmässig durchgeführte Wirkungsmonitoring.
- Rechtliche Grundlage der Aktivitäten der FFG bildet das Bundesgesetz zur Errichtung der Österreichischen Forschungsförderungsgesellschaft mit beschränkter Haftung (Österreichische Forschungsförderungsgesellschaft mbH-Errichtungsgesetz – FFG-G)

AWS – Austria Wirtschaftsservice Gesellschaft

- Anders als die FFG und der FWF handelt es sich bei der AWS *nicht um einen Forschungsintermediär im klassischen Sinne*, sondern um eine Förderbank der öffentlichen Hand. Ihr Aufgabenbereich schliesst sich an die Aufgaben der FFG an und setzt dort ein, wo es in F&E-Projekten um die Entwicklung marktreifer Produkte oder Dienstleistungen geht.
- Der Aufgabenbereich der AWS ist damit keine Programm-Forschungsförderung im engeren Sinne, sondern ist eher im Bereich Unternehmensfinanzierung und -förderung anzusiedeln.
- Dennoch ist die AWS, i.d.R. in enger Kooperation mit der FFG, in der Abwicklung diverser Bundesforschungsprogramme involviert.
- Die Abwicklung von Förder- und Finanzierungsprogrammen erfolgt mithilfe interner Datenbanksysteme, in denen dafür relevante Informationen erfasst, verarbeitet und ausgewertet werden.
- Für die Integration der AWS in eine potenziell zukünftige F(F)DB wäre vorgängig sauber zu trennen, welche Aktivitäten der AWS tatsächlich der Forschungsförderung zuzurechnen sind und welche Daten dieser Aktivitäten erfasst werden müssten.

3.2.4 Empfänger von Fördermitteln

Auch auf Seiten der Empfänger bzw. der empfangenden Institutionen von Fördermitteln werden eine Reihe von Datenbanken unterhalten, die zum einen Daten zu Projekten enthalten, die im Rahmen von Forschungsprogrammen gefördert werden und zum anderen Daten zu Forschungsprojekten, die unabhängig von Programmforschung an der Institution durchgeführt werden. Exemplarisch genannt werden können:

- Die Publikationsdatenbank⁴ sowie die Projektdatenbank (TISS)⁵ der TU Wien
- Das CRIS⁶ der Universität Wien sowie das assoziierte Volltext-Repository u:scholar⁷
- Das Forschungsdokumentationssystem der Universitäten Linz⁸ und Salzburg⁹

4. Anwendungsmöglichkeiten und Ansprüche

4.1 Nutzen und Anwendungsfälle einer F(F)DB

Die in Kapitel 3.2 dargestellte Situation zur Forschungsförderung in Österreich zeigt auf, wie fragmentiert die Datenlandschaft der *Forschungs(förderungs)-informationen* in Österreich ist. Historisch gewachsene Strukturen mit einer Vielzahl von Institutionen mit unterschiedlichen Leistungsaufträgen erfassen, speichern und verwenden Daten, ohne übergeordnete nationale Struktur. Die existierenden Datenbanksysteme sind gezielt für den jeweiligen institutionellen Kontext geschaffen oder angepasst worden. Sie sind fester *Bestandteil etablierter Prozesse* und damit *effiziente Werkzeuge in der Leistungserbringung* dieser Institutionen.

Somit sind auf der einen Seite die bestehenden Lösungen passgenau für ihren jeweiligen Anwendungszweck geschaffen worden und werden in ihren Institutionen effizient eingesetzt. Betrachtet man also das gesamte FTI-System Österreich liegen im Prinzip alle relevanten Informationen über die öffentliche Forschungsförderung bereits vor.

Auf der anderen Seite bedeutet die bestehende Situation aber gleichzeitig, dass es keine übergreifende Struktur gibt, die diese umfassenden, an unterschiedlichen Stellen abgelegten Informationen zusammenführt und nutzbar macht. Das ist das zentrale Problem, dass der Rechnungshof (2016) anspricht, wenn er konstatiert, dass die flächendeckende Erhebung von Forschungsförderungsdaten nicht möglich ist, da eine Gesamtschau nach geförderten Projekten und Förderungsnehmern fehlt.¹⁰ Wenn sich ein interessierter Akteur über öffentlich geförderte Forschungsprojekte in Österreich informieren möchte, ist dies heute nur mit erheblichem Aufwand und unter Einbezug einer Vielzahl von Institutionen möglich. Für die Beschaffung von Detailinformationen

⁴ <https://publik.tuwien.ac.at/>

⁵ <https://tiss.tuwien.ac.at/>

⁶ <http://ucrisinfo.univie.ac.at/>

⁷ <http://openaccess.univie.ac.at/publikationsdienste/uscholar/>

⁸ <https://fodok.jku.at/>

⁹ <http://fodok.uni-salzburg.at/>

¹⁰ Stand 2017 ist anzumerken, dass dies zumindest für die Bundesebene nur bedingt zutrifft, da die Transparenzdatenbank – wenn auch teilweise zeitlich verzögert – einen Grossteil der effektiv abgewickelten Zahlungen der Forschungsintermediäre beinhaltet. Diese Informationen sind jedoch nur bedingt hilfreich. Vgl. dafür Kapitel 3.2.2

über die gesamten in Österreich erbrachten Forschungsleistungen gestaltet sich diese Situation noch komplizierter.¹¹

Der Rechnungshof (2016) moniert vor allem zwei Punkte, die seiner Meinung nach die Prüfung der Einrichtung einer FFDB rechtfertigen: das ausschliessen *unerwünschter Mehrfachförderungen* sowie die *Abbildung der gesamten gebietskörperschaftsübergreifenden Zahlungsströme* in der österreichischen Forschungsförderung.

Es gibt jedoch eine Reihe weiterer Anwendungsfälle, die für andere Anspruchsgruppen des österreichischen FTI-Systems relevant sind und von der Einrichtung einer F(F)DB profitieren könnten. Welchen *Nutzen* die verschiedenen Anspruchsgruppen aus einer F(F)DB ziehen könnten, ist im Folgenden stichwortartig zusammengestellt.

Politik

- Detaillierter Überblick über die für die Forschungsförderung eingesetzten Gelder
→ Politische Steuerungsmöglichkeiten, regelmässige Prüfung und Weiterentwicklung der Wissenschaftspolitik
- Erhöhte Transparenz der öffentlichen Förderung
- Zurverfügungstellung eines gesicherten Datenbestands für interessierte Akteure bei erhöht verfügbarer Datenqualität und -quantität
- Nationale und internationale Analysen des FTI-Systems werden durch einen kontinuierlich aktualisierten und ständig verfügbaren Datenbestand erheblich vereinfacht

Forschungsintermediäre

- Gesamthafter Überblick über öffentliche Programmförderung
- Institutionsübergreifende Auswertungen und Berichterstattungen können vereinfacht und zu einem gewissen Grad automatisiert werden
- „Horizontale“ Nachverfolgung von Projekten und Forschenden (Grundlagenforschung → Marktreife)
- Internationale Vergleichbarkeit – Forschungsförderung kann auf aktuelle Entwicklungen im Ausland abgestimmt werden.
- Anhaltspunkt für die Prüfung möglicherweise ungewollter Mehrfachförderung

Forschende

- Einfacher und transparenter Zugang zu den Daten der öffentlichen Forschungsförderung (z.B. für bibliometrische Forschung)
- Informationsgrundlage für die Positionierung und Konzeption neuer Forschungsgesuche
- Datenbank zur Identifikation möglicher Kooperationspartner und -projekte (Vernetzungsfunktion)
- Recherchewerkzeug für wissenschaftliche Forschung
- Datengrundlage für Ansätze der Netzwerkanalyse

¹¹ Vgl. Kapitel 3.2 für Ausführungen zum Thema Forschungsinformationen und Forschungsförderungsinformationen

Evaluierende

- Einfacher und transparenter Zugang zu den Daten der öffentlichen Forschungsförderung für Evaluationsbedürfnisse¹².

Wirtschaft

- Identifikation möglicher Fachexperten und Kooperationspartner für wissenschaftlich-wirtschaftliche Kooperationsprojekte
- Überblick über den aktuellen Stand der Forschungsförderung als Grundlage für die eigene Innovationstätigkeit

Medien und Journalisten

- Recherchewerkzeug für journalistische Berichterstattung
- Identifikation möglicher Fachexperten und Gesprächspartner

Öffentlichkeit

- Transparenter Zugang zu Ergebnissen und Informationen der öffentlich finanzierten Forschung in Österreich
- Erfüllung des Anspruchs auf öffentliche Rechenschaftspflicht

Es sei an dieser Stelle nochmals explizit erwähnt, dass es erhebliche Unterschiede zwischen einer FFDB und einem nationalen Forschungsinformation (NFIS) gibt. Während eine FFDB ausschliesslich öffentlich geförderte Programmforschungsprojekte aus Datenbeständen der Förderungsgeber/-intermediäre sammelt und zugänglich macht, ist der Anspruch einer nationalen FDB grundsätzlich *alle (soweit dies rechtlich möglich ist)*¹³ Forschungsprojekte staatlich finanzierter Forschungseinrichtungen zugänglich zu machen. Da ein derartiger Ansatz sowohl Forschungsförderungsinformationen als auch generelle Forschungsinformationen enthalten würde, würden die dadurch generierten Nutzen weit über die oben genannten hinausgehen.

4.2 Kerndatensatz

Unabhängig von der gewählten technischen Lösung, von den Inhalten und unabhängig von der regionalen Ebene (national oder EU) auf der ein Datenbanksystem entwickelt wird, ist das zentrale Element eines jeden Datensatzes ein Forschungsprojekt. In welcher Form Forschungsprojekte als hochdimensionales Objekt beschrieben werden und welche Informationen in welchem Format benötigt werden (Datenmodell), hängt in erheblichem Masse davon ab, wofür diese Daten benötigt werden. Eine Hochschule, die Forschungsprojektdatensätze sammelt, um differenziert über ihren Forschungsoutput zu berichten, benötigt andere Informationen als eine Forschungsgruppe, die sich mit

¹² Es ist dabei zu berücksichtigen, dass für tiefgehende oder spezifische Evaluationen von Detailfragen vermutlich auch in Zukunft manuelle ad-hoc Auswertungen der Primärdaten der Forschungsintermediäre notwendig sein werden, da der Datenbedarf für solche Evaluationen häufig über den im Kapitel 4.2 spezifizierten Kerndatensatz hinausgeht

¹³ Es gibt Forschungsprojekte (z.B. mit Verwendung von sensiblen Informationen, oder in Kooperationsprojekten mit der Privatwirtschaft) deren Inhalte oder Existenz mit zusätzlichen Vertraulichkeitsvereinbarungen geschützt sind. Selbstverständlich sind derartig projektspezifische Rechtsgrundlagen bei der Veröffentlichung von Projektinformationen zu berücksichtigen. Prüfung und Berücksichtigung dieser Rechtsgrundlagen erfolgt jedoch bereits bei den Quellinstitutionen und würde gemäss deren Spezifikation auch in der Datenbasis einer F(F)DB berücksichtigt.

Karriereverläufen von Nachwuchswissenschaftlern beschäftigt. In diesem Bericht stehen gemäss Leistungsauftrag nationale Forschungsförderungsinformationen im Vordergrund, unter Berücksichtigung anderer Lösungen, wie z.B. einer EU-weiten Lösung oder einem NFIS.

Die *Definition eines Datenmodells* ist für die Spezifikation einer technischen Lösung von grösster Bedeutung und hat massgeblichen Einfluss auf die Kosten eines Projekts. Für die Konkretisierung der Leistungsansprüche, die an ein derartiges Datenmodell gestellt werden und für die Definition eines effektiv zu erhebenden *Kerndatensatzes*, empfehlen wir in einem nächsten Schritt einen partizipatorischen Prozess einzuleiten, bei dem in enger Abstimmung mit den relevanten Institutionen des FTI-Systems (Ministerien, Intermediäre, Hochschulen, etc.) ein *effektiv zu verwendendes* Rahmenkonzept entwickelt wird. In einem solchen Rahmenkonzept werden die inhaltlichen und technischen Details eines Kerndatensatzes Forschung für Österreich erarbeitet.

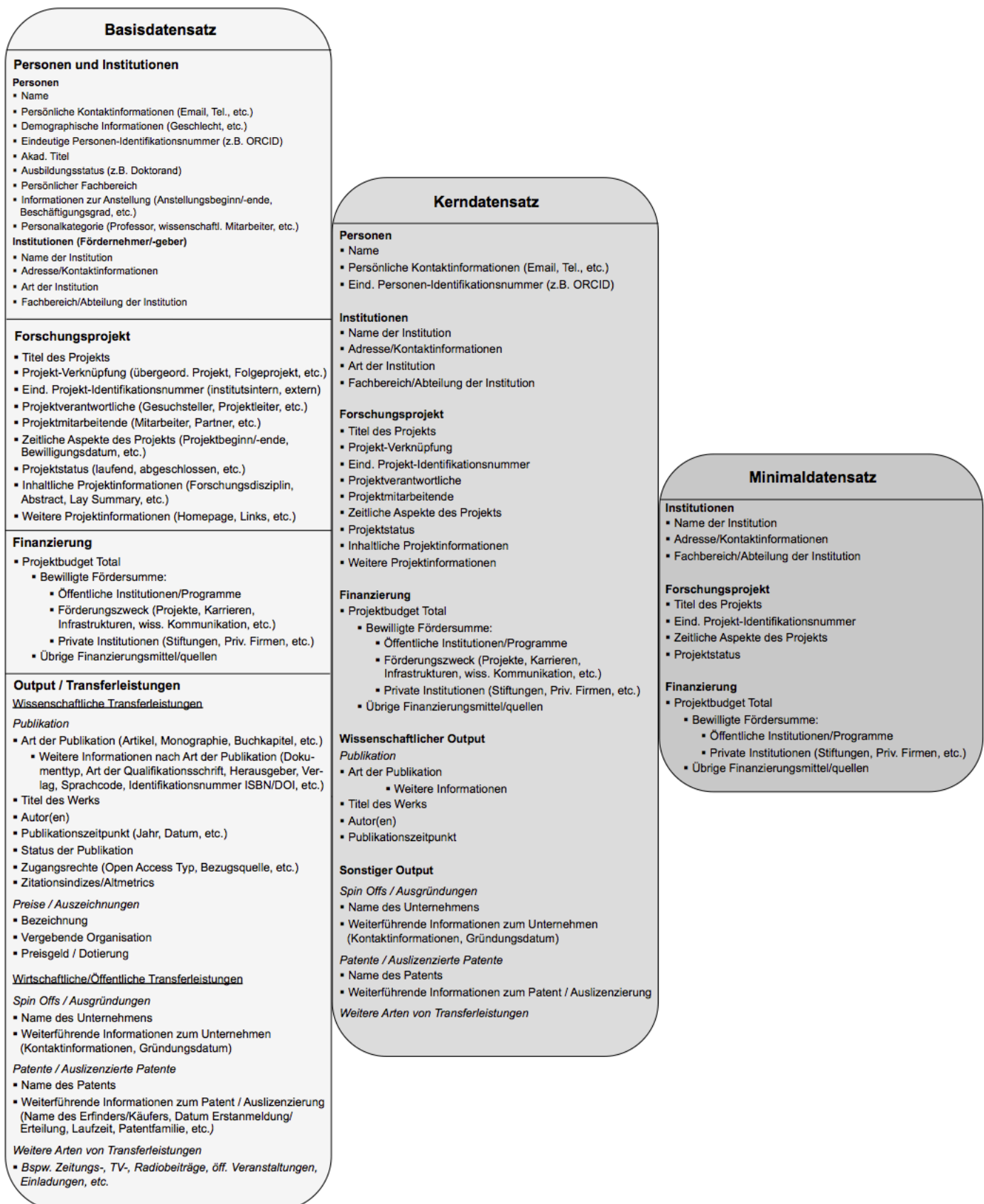
In diesem Kapitel werden exemplarisch Erkenntnisse aus Forschungsprojekten zum Thema Standardisierung von Forschungsinformationen präsentiert und ausgeführt, welche internationalen Standards und Ansätze es zur Standardisierung von Forschungsinformationen gibt. Diese Übersicht kann als Input für den empfohlenen partizipativen Prozess mit den Fachexperten der FTI-Institutionen dienen.

Es lassen sich verschiedene Datenebenen unterscheiden, die einen unterschiedlichen Detail- und Aggregationsgrad aufweisen. Diese Ebenen werden zum einen durch die entsprechenden Leistungsansprüche der verschiedenen Anspruchsgruppen definiert. Zum anderen ergeben sich in Abhängigkeit des Detailierungsgrads offene Fragen hinsichtlich der rechtlichen Rahmenbedingungen, die berücksichtigt werden müssen (vgl. Kapitel 4.2).

Die folgende Tabelle stellt drei verschiedene Varianten dar, die auf der einen Seite verschiedene Informationstiefen aufweisen und auf der anderen Seite unterschiedliche Ansprüche an Datenstruktur, -erhebungen und rechtliche Rahmenbedingungen stellen. Bei der Zusammenstellung dieser Liste wurde zunächst ausschliesslich die inhaltliche Ebene adressiert.

Den höchsten Detailierungsgrad weisen die *Basisdaten* auf. Sie enthalten die wichtigsten Informationen, die für die Beschreibung eines Forschungsprojekts benötigt werden und bieten damit viele Auswertungsmöglichkeiten. Der *Kerndatensatz* ist dagegen eine reduzierte Form des Basisdatensatzes und berücksichtigt datenschutzrechtliche Aspekte sowie bereits existierende Standards und Bewilligungen der Forschungsintermediäre des Bundes. Er entspricht damit am ehesten den Daten, die bereits von den Intermediären auf Bundesebene zur Verfügung gestellt werden. Der *Minimaldatensatz* ist eine nochmals reduzierte Form des Kerndatensatzes, der vollständig ohne personalisierte Daten auskommt und – wie der Name schon sagt – als absolutes Minimalszenario dient, um Forschungsprojektdaten zumindest rudimentär nutzen zu können.

Abbildung 2: Basisdatensatz, Kerndatensatz und Minimaldatensatz im Überblick



Quelle: Rütter Soceco, eigene Darstellung

4.2.1 Datenschutz und rechtliche Fragen

Die folgenden Einschätzung zur rechtlichen Situation der Datenverwendung einer FFDB stützen sich im wesentlichen auf ein zum Thema Forschungsförderungsdatenbank erstelltes Gutachten des Bundeskanzleramtes (Bundeskanzleramt-Verfassungsdienst, 2012), auf Gespräche mit Forschungsintermediären und Bundesämtern sowie auf institutionelle Rechtsgrundlagen (wie z.B. das TDB-Gesetz und das FFG-Gesetz). Es handelt sich demnach ausschliesslich um eine Übersicht zum Thema und in keinsten Weise um rechtsverbindliche Aussagen.

Die zentrale Frage, die sich auf rechtlicher Ebene stellt, ist wie eine FFDB mit *personenbezogenen Daten* umgeht. Wie in Kapitel 4.2 beschrieben, wäre es grundsätzlich möglich, eine FFDB *ohne* personenbezogene Daten zu erstellen. Ein derartiges Szenario würde jedoch die Anwendungsfälle und den Nutzen einer derartigen Datenbank erheblich reduzieren. Geht man also im folgenden davon aus, dass personenbezogene Daten (Namen, Adressen, etc.) Bestandteil eines Forschungsprojekt-Datensatzes sind, dann besteht grundsätzlich Anspruch auf Geheimhaltung dieser Daten (Grundrecht auf Datenschutz, gemäss § 1 Abs. 1 Datenschutzgesetz (DSG)¹⁴).

Es gibt drei Möglichkeiten, wann dieser Schutz nicht gilt.

- Es ist im lebenswichtigen Interesse des Betroffenen (§ 1 Abs. 2 DSG)
- Der Betroffene gibt seine Zustimmung zur Veröffentlichung (ebd.)
- Bei Eingriffen einer staatliche Behörde, „*auf Grund von Gesetzen, die aus den in Art. 8 Abs. 2 der Europäischen Konvention zum Schutze der Menschenrechte und Grundfreiheiten (EMRK), BGBl. Nr. 210/1958¹⁵, genannten Gründen notwendig sind*“ (ebd.)

In bereits bestehenden Datenbanklösungen der öffentlichen Hand, die personenbezogene Forschungsförderungsdaten sammeln und verfügbar machen, werden unterschiedliche Rechtsgrundlagen zu diesem Zweck verwendet. Für die Transparenzdatenbank, welche rigide Authentifizierungsmechanismen und Rollendefinitionen verwendet, wurde bspw. eine eigene Rechtsgrundlage geschaffen, die die Themen Datenverwendung und Datenschutz regelt (vgl. z.B. TDBG 2012, § 2). Die Veröffentlichung von Projektinformationen (inkl. Name und Emailadresse des Projektleiters) im Rahmen des Project Finders der FWF erfolgt auf Basis der *Zustimmung* besagter Personen. Bei jedem Projektantrag wird diese freiwillige und widerrufbare Einwilligung der Projektleitenden eingeholt.

§ 4 Z 14 DSG 2000 spezifiziert, dass derartige Einwilligungen

- ohne Zwang,
- in Kenntnis der Sachlage
- und für den konkreten Fall

erfolgen müssen. Vor allem letzterer Punkt müsste für den Fall einer Weiterverwendung der durch Forschungsintermediäre erhobenen im Rahmen einer FFDB intensiv geprüft werden.

Der Verfassungsdienst des Bundeskanzleramtes weist in seiner Stellungnahme zum Thema v.A. auch darauf hin, dass eine *allfällige Veröffentlichung* von personenbezogenen Daten der Forschungsförderung grundsätzlich geklärt werden müsste, welchem Zweck dies dienlich sein soll (vgl. Kapitel 4) und ob dieser Zwecke nicht auch mit ano-

¹⁴ Bundesrecht konsolidiert: Gesamte Rechtsvorschrift für Datenschutzgesetz 2000, Fassung vom 30.08.2017 <https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=bundesnormen&Gesetzesnummer=10001597>

¹⁵ https://www.ris.bka.gv.at/Dokumente/BgblPdf/1958_210_0/1958_210_0.pdf

nymisierten und aggregierten Daten erfüllt werden kann. Der Europäische Gerichtshof hält z.B. in den Urteilen in den verb. Rs C-92/09 und C-93/09, Schecke und Eifert/Land Hessen fest, „dass die Veröffentlichung personenbezogener Daten aller natürlicher Personen, die Empfänger von bestimmten Förderungen sind, ohne nach einschlägigen Kriterien wie den Zeiträumen, während deren sie solche Beihilfen erhalten haben, der Häufigkeit oder auch Art und Umfang dieser Beihilfen zu unterscheiden, unzulässig ist.“ (BKA, 2012).

Ausserdem ist an dieser Stelle zu erwähnen, dass es Forschungsprojekte gibt, die aus anderen Gründen als der Verwendung von personenbezogenen Daten besondere Schutzbedürfnisse haben (vgl. Fussnote 13). Dies ist z.B. bei der Verwendung von sensiblen Daten oder in besonderen Vertrauensverhältnissen (z.B. Entwicklung von Prototypen) der Fall. Projekte mit besonderen Rechtsgrundlagen hinsichtlich Datenschutz müssen notwendigerweise bei jeder Überlegung einer Datenbanklösung besonders berücksichtigt werden.

Der Mehrwert einer F(F)DB ergibt sich nicht zuletzt auch aus der Verknüpfung von Projekt- und Personendaten. Sollte eine derartige Variante in der Detailausarbeitung eines Umsetzungsszenarios vorgesehen sein, wäre vorgängig zu prüfen, auf welcher Rechtsgrundlage die Sammlung und Verwendung dieser personenbezogenen Daten erfolgen kann. Diese Prüfung ist vor allem deshalb wichtig, weil eine F(F)DB – gemäss den in diesem Bericht erarbeiteten Varianten – diese Daten nicht selbst erhebt, sondern als ein Sekundärnutzer fungiert. Im Rahmen einer derartigen Struktur ist demnach die Frage zu klären, in wie weit bestehende Rechtsgrundlagen der Quellinstitutionen für die Weiterverwendung der Daten in einer F(F)DB gültig sind, allenfalls angepasst werden müssen oder ob eine eigene Rechtsgrundlage für die F(F)DB und angehängte Quellinstitutionen geschaffen werden müsste.

4.2.2 Datenstandardisierung

Die fragmentierte Datensituation in den österreichischen Institutionen der Forschungsförderung ist nicht zuletzt ein Ausdruck historisch gewachsener Aufgabenbereiche und Verantwortlichkeiten, sowie eines kontinuierlichen Ausverhandelns von Zuständigkeiten und Kompetenzen. Stand heute ist daraus ein System entstanden, das – zumindest auf Bundesebene – zwar inhaltlich grösstenteils sinnvolle Trennungen und kaum ungewollte Überlappungen aufweist, jedoch gleichzeitig keine übergeordnete Struktur besitzt, die eine einfache Steuerung und Kontrolle ermöglicht. Anders ausgedrückt: das System, wie es gewachsen ist und heute funktioniert, läuft grundsätzlich gut und eingespielt, würde heute jedoch vermutlich nicht so aufgebaut, wenn man es neu entwerfen würde.

Ein wichtiger Aspekt bei derartig systemischen Überlegungen betrifft die Standardisierung und Harmonisierung von Daten. Heute verwendet quasi jede Institution im FTI-System eigene interne Datenbanksysteme mit eigenen Datenmodellen zur Erfassung, Speicherung und Auswertung von Forschungs(förderungs)informationen bzw. Forschungsprojekten (oder „Förderfällen“) und andere damit verknüpfte Daten. Selbst funktional homogene Institutionsgruppen wie Hochschulen oder Förderungsintermediäre verfügen nicht über aufeinander abgestimmte Datenbanksysteme, die einen einfachen und vernetzten Zugriff ermöglichen würden. Selbstverständlich lassen sich diese verschiedenen Daten mit etwas Aufwand kombinieren und aufeinander abstimmen, alleine weil es sich um grundsätzlich ähnliche Daten handelt (z.B. Publikationen, oder Projektinformationen). Derartige Funktionen sind aber aktuell weder etabliert, noch automatisiert vorgesehen. Möchte sich ein interessierter Akteur heute über die Forschungslandschaft in Österreich informieren, ist dies anhand von statistischen Indikatoren, Jahresberichten und Reports auf quantitativer Ebene gut möglich. Möchte er jedoch inhaltliche

Informationen über die Forschung in Österreich gewinnen, stehen weit über 30 nationale und regionale Datenbanken von Forschungs- und Förderungsinstitutionen zur Verfügung, die zu dieser Frage mit diversen Überschneidungen Auskunft geben können.¹⁶

Vor dem Hintergrund individuell effizient aufgestellter Institutionen, die mit ihren spezifischen Systemen ihre Leistungsaufträge erfüllen, macht es jedoch wenig Sinn neue Standards zu schaffen (oder zu adaptieren) und systemweit einen erheblichen Anpassungsaufwand zu verursachen. Stattdessen können in solchen existierenden Systemen Vernetzungsansätze helfen (vgl. Kapitel 5.3), die *auf bestehende Systeme aufbauen* und in der Lage sind mit hochgradig vernetzten, heterogen strukturierten Daten von verschiedenen Quellen umzugehen. Bei der Konzeption eines solchen Vernetzungssystems sollten jedoch *unbedingt* existierende Standards für die Strukturierung von Forschungsinformationen berücksichtigt werden. Der grosse Vorteil entsteht dann dadurch, dass durch die Kompatibilität des Vernetzungssystems mit internationalen Datenstandards alle dort eingespeisten Quelldaten automatisch ebenfalls im entsprechenden Format abrufbar werden, ohne dass die Quellinstitution ihre Datenbestände verändern muss¹⁷.

CERIF (Common European Research Information Format) ist ein standardisiertes Datenmodell für Forschungsinformationen und einer Empfehlung der EU für die Mitgliedsstaaten. Die Weiterentwicklung von CERIF wurde von der EU an euroCRIS übertragen, einer Non-Profit-Organisation für die Kompatibilität von Forschungsinformationssystemen. (FIS/CRIS). Zusätzlich zu einem Domain-Modell definiert CERIF ein formal-relationales Datenmodell (ERM) sowie ein XML Format zum Datenaustausch. Das OpenAIRE Datenmodell ist CERIF-kompatibel und CERIF XML dient OpenAIRE als Grundlage für die Sammlung und den Import von Metadaten aus Forschungsinformationssystemen.

Quelle: freie Übersetzung nach euroCRIS (2015)

Das wichtigste in Europa verwendete System ist das sog. CERIF System (vgl. Box), welches auch die Grundlage bereits existierender FIS im Rahmen der EU-Forschungsprogramme bildet. Ein weiterer Ansatz kommt von CASRAI (Consortia Advancing Standards in Research Administration Information) einer Non-Profit-Organisation bestehend aus internationalen Forschungs- und Forschungsförderungsinstitutionen. Es gibt andauernde Ansätze eines Mappings zwischen diesen beiden Systeme herzustellen¹⁸. Für die Speicherung und Identifikation von Personendaten sollte die Adaption des ORCID System geprüft werden. Einem internationalen Non-Profit Konsortium dessen Aufgabe die eindeutige Identifikation und Zuordnung von Forschenden weltweit ist. Zahlreiche europäische Länder, Universitäten und Forschungsförderer sind dort bereits als Mitglied aktiv und setzten das System ein. In Österreich ist bisher ausschliesslich der FWF an ORCID beteiligt.

¹⁶ Selbstverständlich gibt es auch die Möglichkeit internationale Publikationsdatenbanken abzufragen, die sich in den letzten Jahren hochgradig verknüpft haben und heute einen sehr breiten Überblick über Forschungsausgaben (Publikationen) liefern.

¹⁷ Ob dieser Vorteil realisiert werden kann, hängt von der gewählten Variante ab (vgl. Kapitel 5).

¹⁸ Vgl. <http://dspacecris.eurocris.org/handle/11366/204>

Aufbauend auf den Status Quo und die skizzierten Anwendungsfälle und Nutzen, die eine Forschungsförderungsdatenbank generieren kann (vgl. Kapitel 4.1), werden im Folgenden verschiedene Varianten skizziert, die für eine Einrichtung einer FFDB denkbar sind. Es werden *drei technische Varianten* unterschieden, die sich in erster Linie in der Art und Weise wie Informationen erfasst, ausgewertet und zugänglich gemacht werden unterscheiden. Die letzte Variante (4) unterscheidet sich von den ersten drei Varianten dahingehend, dass es sich um ein *Querschnittsszenario* handelt. Dies ist so zu verstehen, dass jede der ersten drei Varianten mit der vierten Variante kombiniert werden kann.

Eine übergeordnete Einschätzung der *gesamten* Aufwände und des Nutzens der einzelnen Varianten findet sich am Ende des Kapitels.

5.1 Variante 1: Meta-Projektfinder ohne weitere Datenspeicherung

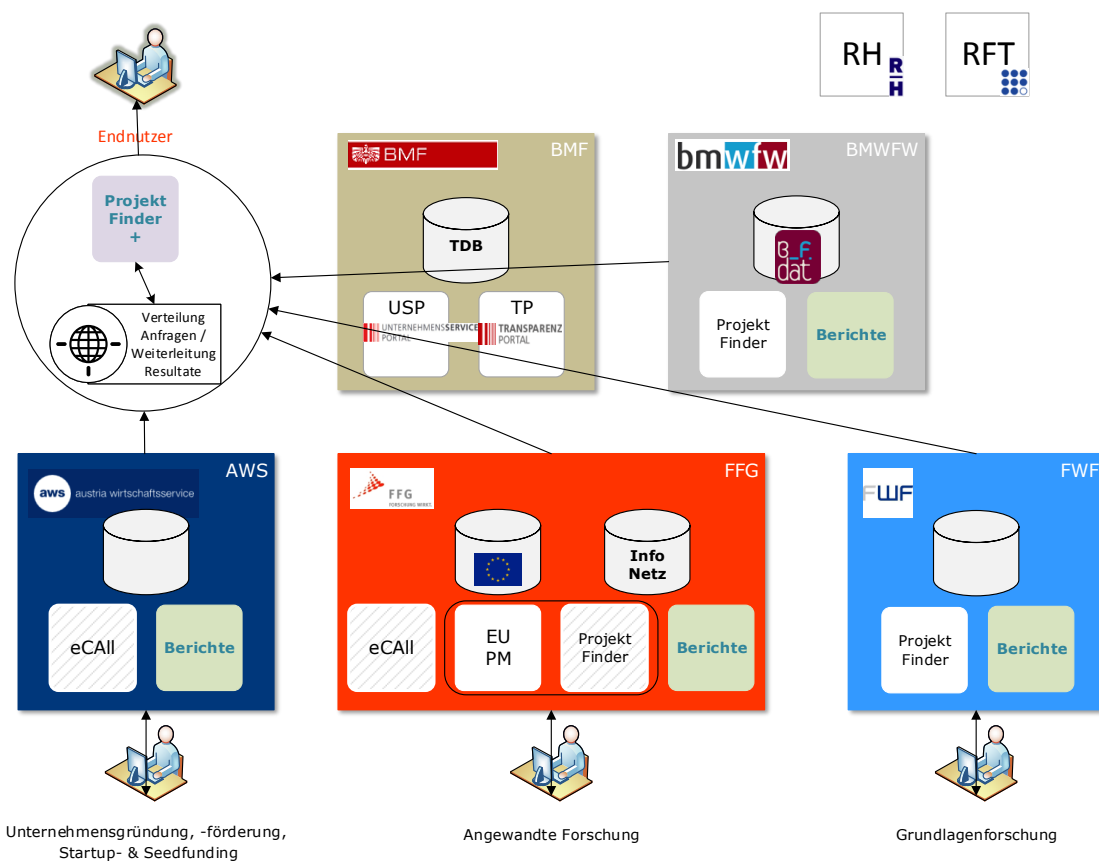
Die Forschungsförderungsinformationen des FWF werden intern in einem Datenbanksystem gespeichert und (in reduzierter Form) über einen internetbasierten Projektfinder der Öffentlichkeit zur Verfügung gestellt. Der Projektfinder zeigt ausschliesslich Informationen¹⁹ zu Projekten, die vom FWF bewilligt wurden. Auch die Bundesforschungsdatenbank bietet eine Art Projektfinder (Webinterface zum Abrufen von Datensätzen). Ein ähnliches Webinterface ist bei der FFG aktuell im Aufbau. Diese verschiedenen Projektfinder dienen der Abfrage von Projekten aus der institutionsinternen Datenbank und sind nicht institutionsübergreifend konzipiert. Sie dienen zum einen als Transparenzwerkzeug der Institutionen und zum anderen bieten sie interessierten Stellen ausserhalb der Forschungsintermediäre einen niedrigschwelligen Zugang zu grundlegenden Projektinformationen.

Eine Variante, um Projektinformationen der drei grossen Forschungsintermediäre und der Bundesforschungsdatenbank zugänglich zu machen, ist der Aufbau eines vergleichbaren Webinterfaces (Projekt-Finder+), das – analog zu den institutionsspezifischen Lösungen – so weit wie möglich auf die bereits bestehenden Datenbestände zugreifen kann und für Suchanfragen aus allen relevanten Datenbanken gespeist wird. Man kann sich diese Variante wie einen *Meta-Projektfinder* vorstellen, der es ermöglicht über ein zentrales Interface auf alle Datenbestände gleichzeitig zuzugreifen. Die Daten müssten bei dieser Variante nicht noch einmal in einer eigenen Datenbank gespeichert werden, sondern es würden die Ressourcen der bestehenden Datenbanken genutzt. Um so einen Meta-Projektfinder einzurichten, müssten Schnittstellen zu den einzelnen Datenbanken (FWF, FFG und AWS sowie B_f.dat) eingerichtet werden, die einen Zugang zu den dort etablierten Daten ermöglichen.

Ein *Nachteil* eines Meta-Projektfinders besteht darin, dass die *vorhandenen Datenbanken* durch die zusätzlichen Suchanfragen über das Webinterface potenziell mehr belastet werden. Dies könnte sich negativ auf die Performance dieser Datenbanken auswirken.

In der nachfolgenden Abbildung wird die eben beschriebene Variante 1 schematisch dargestellt.

¹⁹ In diesem Fall: die interne Projektnummer, den Projektleiter (inkl. Emailadresse), die Kuratoriumssitzung zum Projekt, die (primäre) Forschungsstätte inkl. Institut, einen Link zur Internetseite des Projekt(leiters), die Projekt-Laufzeit, die Bewilligungssumme (für neuere Projekte), den aktuellen Projektstatus, einen Abstract, einen Kurzbericht, die Zuordnung zu den Wissenschaftsdisziplinen gemäss ÖFOS Statistik Austria sowie Keywords für das Projekt. Es sind jedoch nicht alle Informationen für jedes Projekt abrufbar.

Abbildung 4 Schematische Darstellung Meta-Projektfinder ohne eigene Datenspeicherung

Quelle: Rütter Sococo, Quantum Analytics, eigene Darstellung

5.2 Variante 2: Ausbau Transparenzdatenbank

Die Transparenzdatenbank (vgl. Kapitel 3.2) enthält – neben einer Reihe anderer Daten zu öffentlichen Leistungsangeboten in Österreich – hochgradig abstrahierte Daten aus dem Forschungsförderungssystem. Diese Daten werden direkt von den Forschungsintermediären bezogen und sind somit zu ihren Datenbeständen redundant. Die Transparenzdatenbank enthält also im Bereich der Forschungsförderung keine Daten, die nicht auch den Forschungsintermediären vorliegen.

Grundsätzlich hat jede natürliche Person, andere juristische Personen (z.B. Firmen) sowie die Institutionen der öffentlichen Hand Zugriff auf die Transparenzdatenbank. Hierbei ist wichtig zu berücksichtigen, dass die Menge an einsehbaren Informationen auf Basis eines Rollensystems strikt geregelt ist. Der Informationsgehalt der Transparenzdatenbank ist in Bezug auf Forschungsförderungsdaten daher sehr gering und die Daten aus der Transparenzdatenbank sind für die Zwecke einer FFDB (vgl. Kapitel 4) somit wesentlich weniger aussagekräftig wie die Daten der Intermediäre aus dem Forschungsförderungssystem selbst. Gleichzeitig bietet die Datenbank jedoch den Vorteil, dass Datenerhebungsprozesse, Schnittstellen, Wartungsprozesse und die gesetzlichen Grundlagen zur Einspeisung der Daten und Nutzung der Datenbank bereits etabliert sind.

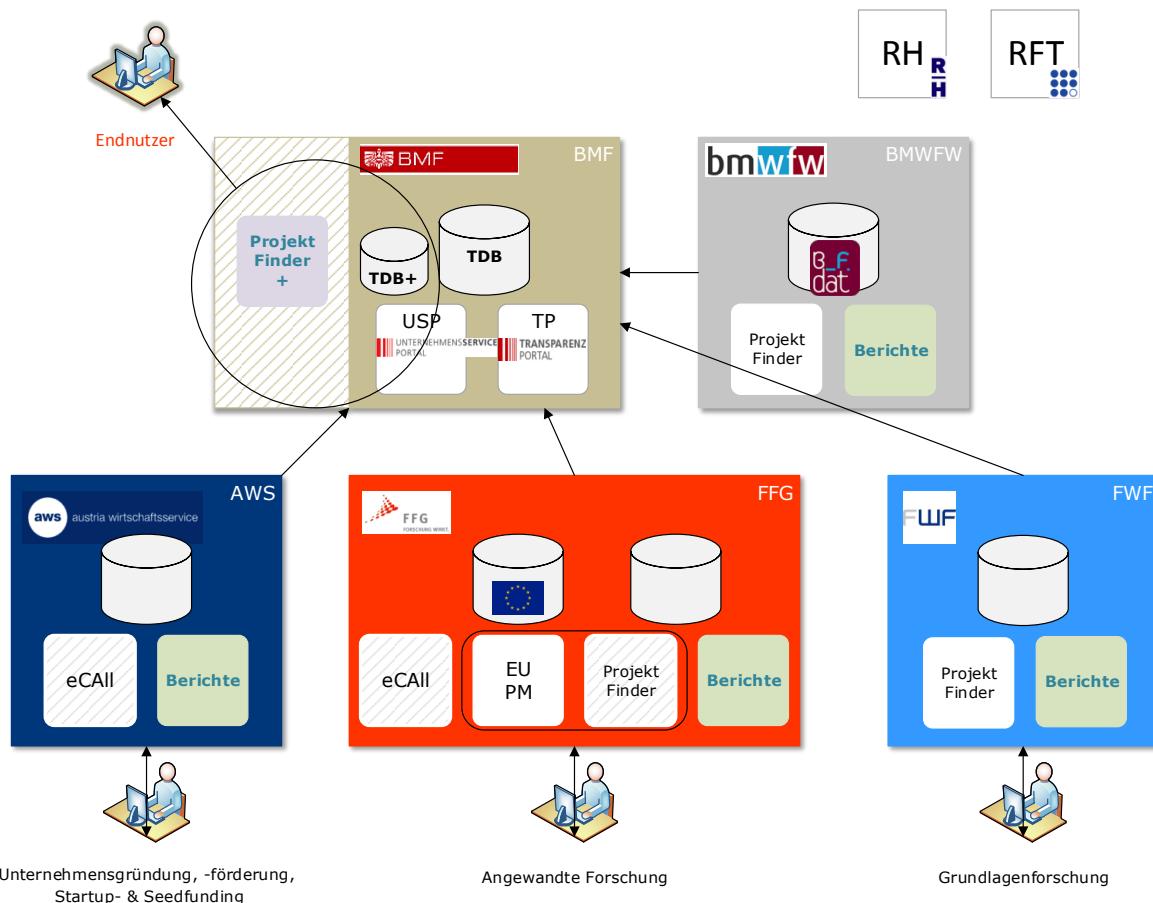
Ein *Ausbau der Transparenzdatenbank* könnte erfolgen, indem man *den Inhalt* der Datenbank mit Informationen aus den Datenbanken von FWF, FFG und AWS sowie der B_f.dat anreichert, z.B. mit sogenannten Meta-Feldern, also Feldern, die das anreichern

der Kerninformation um weitere generische Daten erlauben. Dies würde es ermöglichen, neben den bereits existierenden Daten weitere Informationen aus dem Kerndatensatz (vgl. Kapitel 4.2) in die Datenbank aufzunehmen. Die bereits existierenden Daten umfassen Informationen bzgl. welche Zahlungen erfolgten durch welche Institution und wann mit welchem Betreff an welchen Empfänger.

In diesem Fall wäre es notwendig, die existierenden *Datenaustauschprozesse* und *Schnittstellen* so zu *erweitern*, dass die im Kapitel 4.2 spezifizierten Informationen automatisiert und zeitnah in die Transparenzdatenbank übertragen werden können. Ausserdem müssten die *Zugangsberechtigungen* angepasst werden, damit die Informationen über geförderte Projekte den intendierten Personengruppen zugänglich gemacht werden können. Auch dieses Szenario nutzt als Benutzerschnittstelle ein bereits existierendes Webinterface, das den Zugang zu den in der Datenbank hinterlegten Informationen ermöglicht. Dieses Webinterface müsste so angepasst werden, dass die neu hinterlegten Informationen der Meta-Felder ebenfalls berücksichtigt sind.

Der Nachteil dieser Option ist, dass bereits bestehende, über die Zeit gewachsene Prozesse verändert oder sogar die Datenbankarchitektur komplett überarbeitet werden müsste. Aus diesem Grund wäre es vermutlich einfacher, eine von Grund auf neue Datenbank speziell für diesen Zweck zu errichten (vgl. Variante 3).

Abbildung 5 Schematische Darstellung zum Ausbau der Transparenzdatenbank



Quelle: Rütter Sococo, Quantum Analytics, eigene Darstellung

5.3 Variante 3: Aufbau einer neuen Graphdatenbank

Eine weitere Möglichkeit ist der Aufbau einer *neuen Datenbank* mit den Daten von FWF, FFG, AWS²⁰ und B_f.dat. Hierfür ist es notwendig, *Schnittstellen* zu den zugrundeliegenden Systemen zu erstellen und eine *übergeordnete Datenbankarchitektur* bzw. ein Datenmodell zu entwickeln, sodass nichtstandardisierte Daten in einer Logik zusammengefasst werden, die eine einfache und intuitive Abfrage erlaubt.

Diese Option leidet nicht unter dem Nachteil einer potenziellen Mehr-Belastung vorhandener Datenbanken, da nur zu bestimmten Zeitpunkten (z.B. nachts) Daten aus den zugrundeliegenden Datenbanken in die neue Datenbank übertragen werden. Auch würde die neue Datenbank von Anfang an für den entsprechenden Zweck konzipiert und aufgebaut. Eine neukonzipierte Datenbank würde daher nicht unter dem in Variante 2 beschriebenen Nachteil leiden, dass ein bestehendes System mitsamt Prozessen verändert werden müsste.

Nachteil der hier beschriebenen Variante 3 ist jedoch, dass die neue Datenbank eigene Ressourcen bräuchte, da die Daten physisch gespeichert werden müssten. Im Vergleich zur Variante ohne physische Datenspeicherung ist daher mit einem erhöhten Wartungsaufwand zu rechnen.

Als Technologie wird hier eine sogenannte Graph- oder Netzwerkdatenbank vorgeschlagen. Diese speichert statt ganze Tabellen pro Entität (z.B. Person, Projekt, Interaktion, etc.), jeweils nur eine Instanz einer Entität (z.B. «Johann», «Projekt

„Basic principle of CERIF: most of the characteristics (attributes) of an object (entity) are not stored with the entity (in the entity table) but expressed through „linking entities“ (in database terms: linking tables), allowing multiple roles/characteristics to be expressed for the same aspect. Only the absolute unique characteristics of an entity are stored in the entity table.“

Quelle: (euroCRIS, n.d.)

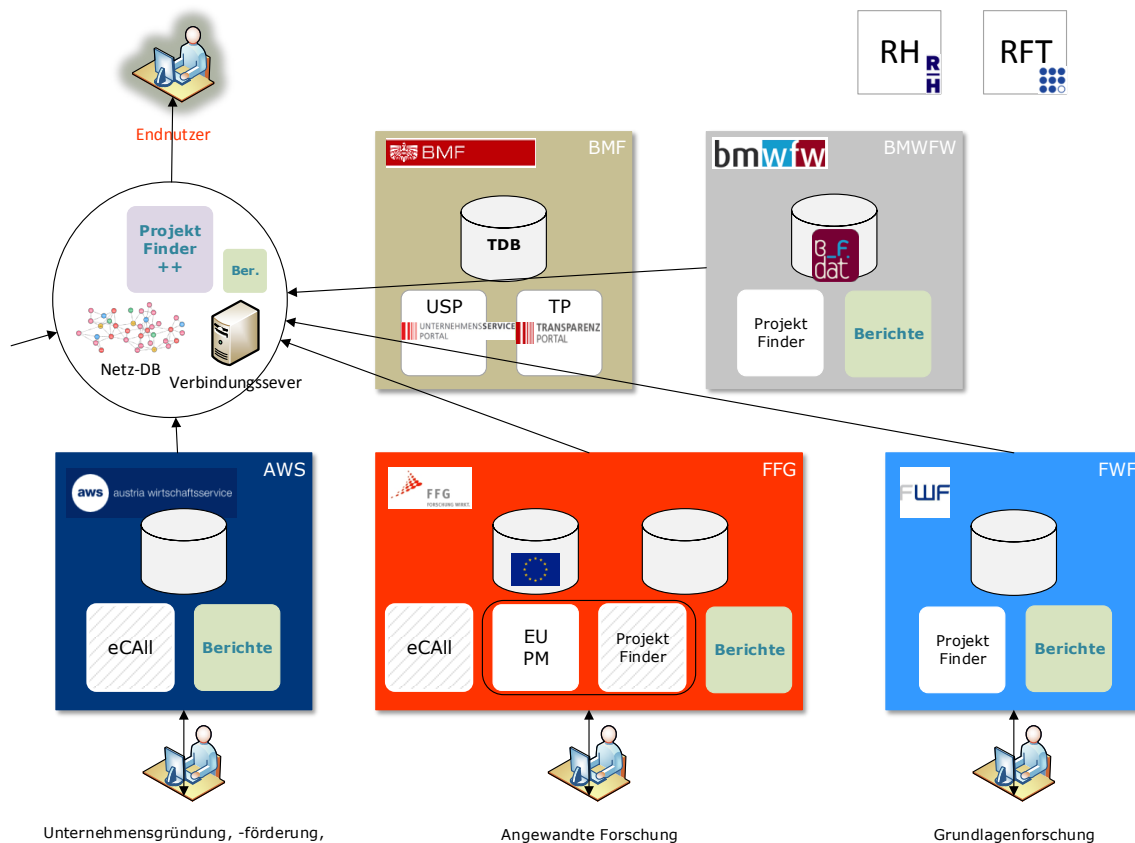
Nanostrukturen», «Email 12.01.2015 10:04:22», etc.) zusammen mit explizit allen bestehenden Verknüpfungen zwischen den Entitäten. Diese Speichermethodik entspricht in hohem Masse dem Grundsatz des CERIF Systems (vgl. Box) und würde eine entsprechende Standardisierung erheblich vereinfachen.

Solche Graphdatenbanken erlauben sehr effiziente Abfragen für komplexe, vernetzte Fragestellungen. Sie können zudem einfach und in sehr kleinen Schritten von beliebiger Art und Struktur (im Extremfall einzelne Entität-Instanzen) erweitert werden. Sie sind sozusagen «schema-frei». Die traditionellen relationalen Datenbanken (mit Tabellen) brauchen dagegen – weil sie mehr Struktur verlangen («schema-pflichtig») – bei Erweiterungen häufig aufwändige strukturelle Anpassungen. Diese Anpassungen sind teuer, langsam und störend.

Die Graphdatenbank erlaubt also eine inkrementelle Integration von Informationen aus vielen unterschiedlichen Quellen. Gleichzeitig bietet sie sehr effiziente Abfragemöglichkeiten für Fragestellungen zu Themen, bei denen Vernetzung eine zentrale Rolle spielt. Daraus resultiert ein verbessertes Webinterface zur Projektsuche, das in der Lage ist, auch komplexe Abfragen effizient und intuitiv darzustellen.

²⁰ Wie in Kapitel 3 ausgeführt, müsste aufgrund der speziellen Natur der AWS zunächst eine Zuordnung der Aktivitäten erfolgen.

Abbildung 6: Schematische Darstellung zum Aufbau einer neuen Graphdatenbank



Quelle: Rütter Sococo, Quantum Analytics, eigene Darstellung

5.4 Variante 4: EU-weite Vernetzung

Die Literaturanalyse und die internationalen Fallstudien haben gezeigt, dass auch andere Länder in Europa und weltweit mit der gleichen Herausforderung konfrontiert sind, wie es aktuell in Österreich der Fall ist. Es gibt verschiedene historisch gewachsene Institutionen, die zu unterschiedlichen Zwecken und mit unterschiedlichen Inhalten einzelne Teile der öffentlichen Forschungsförderung verwalten. Gleichzeitig gibt es eine ganze Reihe von Institutionen auf allen

OpenAIRE (Open Access Infrastructure for Research in Europe) ist eine EU finanzierte Initiative, die sich der Implementierung der Open Access Policy der EU Kommission verschrieben hat. Sie sammelt Publikationen aus institutionellen Repositorien (auch aus themenspezifischen, wie z.B. Europe PMC und arXiv) und verknüpft damit Publikationen mit Forschungs- und Projektinformationen, Datensätzen und Autoreninformationen. Ihr Ziel ist es möglichst viel EU finanzierte Forschung öffentlich zugänglich zu machen.

Quelle: Freie Übersetzung nach European Research Council (n.d.)

Ebenen der nationalen FTI-Systeme, die Forschungsförderungsinformationen und Forschungsinformationen in Datenbanken speichern, auswerten und teilweise öffentlich verfügbar machen. Gleichzeitig stellen alle befragten Institutionsrepräsentanten – vor dem Hintergrund der zunehmenden Internationalisierung des Wissenschaftssystems – den Mehrwert einer *nationalen* ausschliesslich auf Forschungsförderungsinformationen fokussierten Datenbanklösung in Frage (vgl. Kapitel 6). Es wurde im Rahmen der Lite-

raturrecherche auch international nur vereinzelt Ansätze gefunden die eine ausschliesslich auf Forschungsförderungsinformationen fokussierte Lösung verfolgen.

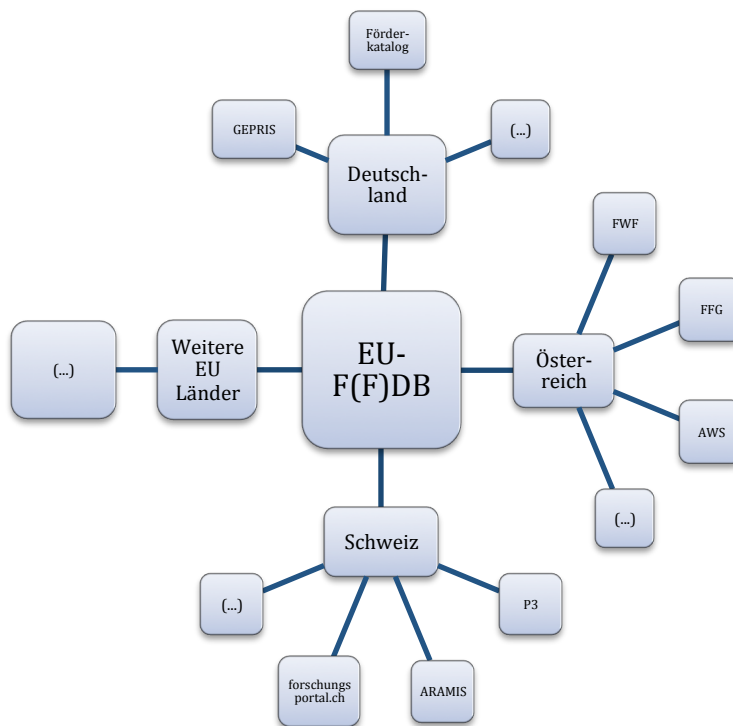
Einen signifikanten Mehrwert, sahen die befragten Personen dagegen in der Idee eines *international abgestützten Forschungsinformationssystems*, welches standardisierte, länderübergreifende Forschungs(förderungs)informationen zur Verfügung stellt (vgl. Kapitel 6). Erhebliche Synergieeffekte werden vor dem Hintergrund der im Rahmen des europäischen Forschungs- und Innovationsprogramms 2007-2013 (FP7) erarbeiteten Grundlagen für die Nutzung von Open-Access Daten im Bereich der wissenschaftlichen Forschung gesehen (vgl. FP7 Clause 39²¹, Horizon 2020 Article 29.2 und 29.3²² sowie den „Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020“²³) sowie in den entsprechenden nationalen Regelungen, die in eine ähnliche Richtung gehen. Hinzu kommt, dass es – primär für die auf EU-Ebene finanzierten Programme – bereits einige Datenbanken gibt, die in diese Richtung gehen (z.B. OpenAIRE, vgl. Infobox). Auch privatwirtschaftliche Akteure sind in diesem Bereich bereits aktiv. So offeriert die Firma ÜberResearch den kommerziellen Zugang zu Informationen von «200 Forschungsförderungsinstitutionen und über 900 Mrd. USD an historischen Projektzusagen» für Universitäten, Stiftungen, Forschende und andere interessierte Personengruppen (ÜberResearch, n.d.).

Wie eingangs des Kapitels 5 erwähnt, handelt es sich bei der hier skizzierten Variante 4 nicht um eine technische Alternative zu den Varianten 1 bis 3, sondern um eine Querschnittsvariante, die grundsätzlich mit jeder der *bisherigen Varianten kombinierbar* ist. Das Ziel der hier beschriebenen Variante ist es, die einzelnen auf nationaler Ebene entstanden Lösungen auf EU-Level zu koordinieren, zu standardisieren und die Daten auf *gesamteuropäischer Ebene* nutzbar zu machen. Alle EU-Staaten (plus möglicherweise die Schweiz und Norwegen), könnten sich dazu verpflichten, einen CERIF kompatiblen Kerndatensatz an Forschungs(förderungs)informationen zur Verfügung zu stellen. Ein solcher CERIF-Datensatz wird zunächst auf nationaler Ebene konsolidiert und anschliessend EU-weit zusammengefasst.

²¹ Vgl. http://ec.europa.eu/research/participants/data/ref/fp7/92570/fp7-lga-clauses_en.pdf

²² Vgl. http://ec.europa.eu/research/participants/data/ref/h2020/mga/erc/h2020-mga-erc-multi_en.pdf

²³ Vgl. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

Abbildung 7 Schematische Darstellung einer EU-weit konsolidierten F(F)DB

Quelle: Rütter Soceco, Quantum Analytics, eigene Darstellung

Jede der drei aufgezeigten Varianten wäre *grundsätzlich* in der Lage, die notwendigen Kerndatensätze auf nationaler Ebene in einer Form zusammenzustellen, sodass sie in einer europäischen F(F)DB verarbeitet werden könnten. Die Komplexität dieser Variante ist wesentlich höher, als bei den anderen Varianten, da es neben dem Aufbau nationaler Lösungen – der teilweise schon abgeschlossen ist – einen EU-weiten Abstimmungsprozess brauchen würde und eine auf EU-Ebene angesetzte Institution, die den Aufbau und den Betrieb einer solchen Datenbank übernimmt. Es wäre denkbar und ggf. sinnvoll, Synergien mit bereits bestehenden Repositorien zu prüfen, da eine Integration in ein bestehendes System ebenso denkbar ist, wie eine Neuentwicklung.

Gleichzeitig wäre der Nutzen sowohl auf europäischer als auch auf nationaler Ebene *erheblich* grösser, als dies bei nationalen Lösungen der Fall wäre. Dies gilt nicht nur für den politischen Steuerungsnutzen, sondern auch für den Nutzen für das FTI-System im Allgemeinen (vgl. Kapitel 4).

Als konkretes Beispiel für eine der Variante 4 ähnlichen EU-weiten Lösung ist die Vereinbarung und anschließende Entwicklung um die EU-Verordnung REMIT, Regulation 1227/2011 – EU-Verordnung über die Integrität und Transparenz des Energiegrosshandelsmarkts, zu erwähnen.

Sie verpflichtet Marktteilnehmer des europäischen Energiemarkts zu verschiedenen Datenlieferungen: Transaktionsdaten mit Lieferort EU, Fundamentaldaten (Kapazität und Nutzung von Anlagen zur Erzeugung, Speicherung, Übertragung von Strom oder Erdgas), Daten zur geplanten oder ungeplanten Nichtverfügbarkeit dieser Anlagen und Daten zur Registrierung beim jeweilig zuständigen EU-Regulator.

Aus dieser Initiative ist u.a. die ENTSO-E Transparency Plattform²⁴ entstanden, eine Datenbank, die diese national aggregierten Daten sammelt und zur Verfügung stellt. Die

²⁴ <https://transparency.entsoe.eu/>

Plattform bietet eine Web-Schnittstelle (restful API) an, über die autorisierte Benutzer eine breite Palette an Informationen abrufen können.

Wie derartig aggregierte Daten im Anschluss ausgewertet werden, zeigt www.electricitymap.org. Diese Webseite bietet auf Basis national gesammelter und europaweit aggregierter Daten eine integrierte, interaktive Sicht auf die Elektrizitätsversorgung für mehrere EU-Länder an. Eine ähnliche Lösung mit national konsolidierten Datenhubs die auf EU-Ebene konsolidiert zusammengefasst werden wäre auch für den Bereich der Forschungs(förderungs)informationen denkbar.

5.5 Die Varianten im Vergleich

Da es sich bei dieser Untersuchung um eine erste Annäherung an das Thema handelt und die skizzierten Varianten als Konzepte erstellt wurden, sind viele Details – sowohl auf technischer als auch auf organisatorischer Ebene – noch offen und bedürfen weiterer Abklärung bzw. Entscheidungen im Projektverlauf. Aus diesem Grund ist die Bewertung der Aufwände der einzelnen Varianten zum jetzigen Zeitpunkt schwierig und mit erheblichen Unsicherheiten verbunden. Um dennoch vergleichende Aussagen über die Varianten machen zu können, wird in diesem Kapitel skizziert, mit welchen inhaltlichen Aufwänden zu rechnen ist, welche Faktoren diese Aufwände massgeblich beeinflussen und wie sich die Varianten hinsichtlich Aufwand und Nutzen voneinander unterscheiden.

Aufwand

Systeme deren primäre Aufgabe in der Datenabfrage und -auswertung liegt und die ihren Informationsinhalt über automatisierte Prozesse aus Quellsystemen beziehen (sog. Auswertungssysteme), sind in der Regel weniger komplex im Aufbau und damit günstiger zu realisieren als Systeme, welche primär für die Ersterfassung von Informationen dienen. Dies liegt vor allem daran, dass Erfassungssysteme zum einen relativ komplexe interaktive Funktionalitäten benötigen, um die vielen Anwenderszenarien zu unterstützen und zum anderen, dass Erfassungssysteme häufig mit institutionellen Prozessen verknüpft sind, die enge Abstimmung benötigen (z.B. Antragsprüfung, Datenverifizierung, etc.). Bei den skizzierten Varianten für eine F(F)DB handelt es sich um ein Abfrage- und Auswertungssystem. Die bereits bestehenden Systeme bei den in Kapitel 3 skizzierten Institutionen sind dagegen Erfassungssysteme mit Abfragemöglichkeiten.

Für jede der Varianten ist zwischen verschiedenen Arten von Aufwand zu unterscheiden. Zum einen treten bei einem Vorhaben wie der F(F)DB politische Aufwände auf. Dazu gehören beispielsweise die *Schaffung gesetzlicher Grundlagen* und die damit verbundenen *parlamentarischen Prozesse*, die *Finanzierung* sowie *Konsultations- und Partizipationsprozesse* mit involvierten bzw. betroffenen Institutionen. Zum anderen fallen *technische Aufwände* an, zu denen zählen bspw. Kosten für die Projektabwicklung, die Datenbankentwicklung, einen Betatest sowie für ein abschliessendes Deployment und für anschliessende Weiterentwicklungsprozesse. Da die notwendigen politischen Aufwände eng mit der letztlich gewählten technischen Detaillösung zusammenhängen, lassen sich diese zum aktuellen Zeitpunkt nicht plausibel abschätzen. Stattdessen wird der Schwerpunkt in dieser Aufwandsschätzung zunächst auf die technischen Aufwände gelegt.

Grundsätzlich unterscheiden sich die technischen Aufwände für ein derartiges Projekt in folgende Kostenarten:

Einmalige Kosten für den Aufbau des Systems

- Projektmanagement
- Einarbeitung und Verständnis bestehender Anwendungsfälle und Informationsmodelle
- Aufbau von Datenmodell/-architektur (Variante 3) bzw. Adaption (Variante 1 und 3)
- Konzeption und Aufbau der benötigten IT-Infrastruktur (eigene oder Cloud-Lösung)
- Etablierung der automatisierten Datenschnittstellen (via API oder direkt, je nach Verfügbarkeit)
- Entwicklung der Benutzeroberfläche (Menschen) und API (Maschine-2-Maschine Communication, v.a. für Variante 4 zwingend)
- Testing & Deployment
- Dokumentierung/Supportprozesse/Schulungen

Laufende Kosten für den Betrieb des Systems

- Hosting
- Wartung, Support & technisches Monitoring (Performance)
- Weiterentwicklung (z.B. durch inhaltliche Evaluation der Abfragen)

In Tabelle 2 werden die Kostenfaktoren pro Variante detailliert aufgeführt, und individuell kommentiert.

Tabelle 2: Detailkommentierung einzelner Arbeitsschritte pro Variante

Komponente	Ansatz	Kommentar
Abfrageprozess	Webinterface	Zentrale Abfrage wird auf allen Quellsystemen individuell abgesetzt
	Ausbau Transparenzdatenbank	SQL-Abfragen werden auf die erweiterte TDB abgesetzt
	Aufbau einer neuen Graphdatenbank	Cypher-Abfragen werden auf die Graphdatenbank abgesetzt
	EU-weite Vernetzung	Cypher-Abfragen werden auf die Graphdatenbank abgesetzt
Aktualisierung	Webinterface	Immer aktuell
	Ausbau Transparenzdatenbank	Regelmässiger (z.B. täglicher) Batch-Import aus allen an das System angeschlossenen Quellen
	Aufbau einer neuen Graphdatenbank	Regelmässiger (z.B. täglicher) Batch-Import aus allen an das System angeschlossenen Quellen
	EU-weite Vernetzung	Regelmässiger (z.B. täglicher) Batch-Import aus allen an das System angeschlossenen Quellen
Anpassung der Quellen	Webinterface	Nicht nötig
	Ausbau Transparenzdatenbank	Nicht nötig
	Aufbau einer neuen Graphdatenbank	Nicht nötig
	EU-weite Vernetzung	Alle EU Teilnehmer müssen eine Standard-Informationspalette zur Verfügung stellen
Authentifizierung	Webinterface	Kann relativ einfach an die Quellen weitergegeben werden
	Ausbau Transparenzdatenbank	Existiert bereits, kann je nach Anwendungsfall erweitert werden (z.B. öffentlich zugänglich)
	Aufbau einer neuen Graphdatenbank	Grundsätzlich konsistent mit der Quelle, je nach Anwendungsfall erweiterbar
	EU-weite Vernetzung	Grundsätzlich konsistent mit der Quelle, je nach Anwendungsfall erweiterbar
Belastung der Quelle	Webinterface	Jede Abfrage belastet mindestens eine Quelle
	Ausbau Transparenzdatenbank	Jede Abfrage belastet die TDB direkt; die Aktualisierung der Quellen passiert in der Nacht
	Aufbau einer neuen Graphdatenbank	Die Aktualisierung der Quellen passiert in der Nacht
	EU-weite Vernetzung	Die Aktualisierung der Quellen passiert in der Nacht
Benutzeroberfläche	Webinterface	Zentrales Portal, ähnlich wie Projekt-Finder
	Ausbau Transparenzdatenbank	Kann durch lokale Datenhaltung effiziente Transformations- und Anreicherungsprozesse ermöglichen
	Aufbau einer neuen Graphdatenbank	Kann durch lokale Datenhaltung effiziente Transformations- und Anreicherungsprozesse ermöglichen
	EU-weite Vernetzung	Kann durch lokale Datenhaltung effiziente Transformations- und Anreicherungsprozesse ermöglichen
Bewilligungen	Webinterface	Bewilligung für die Weiterveröffentlichung muss auf institutioneller Ebene eingeholt werden
	Ausbau Transparenzdatenbank	Existierende Bewilligungen müssen je nach Anwendungsfall angepasst werden
	Aufbau einer neuen Graphdatenbank	Bewilligung für die Weiterveröffentlichung muss auf institutioneller Ebene eingeholt werden
	EU-weite Vernetzung	EU-weite verbindliche Standards müssen national umgesetzt werden und eine Verwendung auf EU-Ebene erlaubt sein
Datenhaltung	Webinterface	Lokal in den Quellen
	Ausbau Transparenzdatenbank	TDB-Schema wird erweitert mit lokal gespeicherten Daten
	Aufbau einer neuen Graphdatenbank	Lokal kompakt und flexibel gespeicherte Daten
	EU-weite Vernetzung	Lokal kompakt und flexibel gespeicherte Daten
Datenmodell	Webinterface	Zentrale Abfragen werden auf lokalen Datenmodellen abgebildet
	Ausbau Transparenzdatenbank	Bestehende und neue Informationen aus allen Quellen müssen innerhalb der TDB harmonisiert werden
	Aufbau einer neuen Graphdatenbank	Muss entwickelt werden; umfasst Informationen aus allen Quellen und kann sehr einfach angepasst oder erweitert werden
	EU-weite Vernetzung	Muss entwickelt werden; umfasst Informationen aus allen Quellen und kann sehr einfach angepasst oder erweitert werden
Hosting	Webinterface	Verbindungsserver bei Drittpartei
	Ausbau Transparenzdatenbank	TDB-Erweiterung, Infrastruktur des BMF
	Aufbau einer neuen Graphdatenbank	Eigene Infrastruktur oder bei Drittpartei
	EU-weite Vernetzung	Eigene Infrastruktur oder bei Drittpartei
Inhalt	Webinterface	Besteht grundsätzlich
	Ausbau Transparenzdatenbank	Besteht grundsätzlich; kann je nach Bedarf erweitert werden
	Aufbau einer neuen Graphdatenbank	Besteht grundsätzlich; kann je nach Bedarf erweitert werden
	EU-weite Vernetzung	Besteht grundsätzlich; kann je nach Bedarf erweitert werden
Kommunikation	Webinterface	Vereinbarung mit Quellenhaltern
	Ausbau Transparenzdatenbank	Vereinbarung mit Quellenhaltern
	Aufbau einer neuen Graphdatenbank	Vereinbarung mit Quellenhaltern
	EU-weite Vernetzung	Vereinbarung mit Quellenhaltern
Performance	Webinterface	Ähnlich wie Quelle; verlangt gut Netzanbindung
	Ausbau Transparenzdatenbank	Würde bei erheblichem Ausbau der Informationsbreite spezifische Adressierung der Datenbankperformance erfordern
	Aufbau einer neuen Graphdatenbank	Speziell konzipiert für komplexe, hochgradig vernetzte Abfragen
	EU-weite Vernetzung	Speziell konzipiert für komplexe, hochgradig vernetzte Abfragen
Schnittstelle auf	Webinterface	Web
	Ausbau Transparenzdatenbank	Web-API oder File-Transfer
	Aufbau einer neuen Graphdatenbank	Web-API oder File-Transfer
	EU-weite Vernetzung	Ergänzend zur standardisierten Informationspalette: erfordert EU-weit standardisierte Datenschnittstelle
Systemsprache	Webinterface	Deutsch, mit mehrsprachigen Inhalten erweiterbar
	Ausbau Transparenzdatenbank	Deutsch, mit mehrsprachigen Inhalten erweiterbar
	Aufbau einer neuen Graphdatenbank	Deutsch, mit mehrsprachigen Inhalten erweiterbar
	EU-weite Vernetzung	Mehrere Sprachen
Wartung	Webinterface	Help-Desk und Support
	Ausbau Transparenzdatenbank	Bestehender Help-Desk und Support
	Aufbau einer neuen Graphdatenbank	Help-Desk und Support
	EU-weite Vernetzung	Help-Desk und Support

Quelle: Rütter Sococo, Quantum Analytics, eigene Darstellung

Die *Kostenbewertungen* reflektieren eine *grobe Einschätzung* der zu erwartenden Aufwände auf Basis der aktuell verfügbaren Informationen und von Erfahrungen in vergleichbaren Projekten. Da die effektiven Kosten in erheblichem Masse von den tatsächlichen Spezifikationen des Leistungsauftrags und von der Struktur der zugrundeliegenden Datenquellen abhängen, lässt sich eine spezifischere Kostenschätzung ohne ein explorativ orientiertes Vorprojekt (mit einem effektiven Datenbezug der involvierten Datenquellen) und eine präzisere Leistungsdefinition nicht durchführen. Um diese Unsicherheit abzubilden, haben wir ein fiktives Szenario kreiert, das einen minimalen Funktionsumfang beinhaltet, der für die Variante 1 und 3 relevant ist und in Form von Kostenbandbreiten berechnet wird. Dieses fiktive Szenario beinhaltet ausschliesslich

- die Integration der drei grossen Forschungsintermediäre (FWF, FFG und AWS), pro weitere Institution ist mit etwa 14-18 zusätzlichen Tagen zu rechnen,
- etwa 15-20 Entitäten im Datenmodell (vgl. Kerndatensatz),
- den Bezug sämtlicher verfügbarer historischer Datensätze gemäss der institutionell vorgesehenen Publikationsrichtlinien,
- die Nutzung einer Cloudlösung für die IT-Infrastruktur,
- die Entwicklung einer Abfrageschnittstelle ohne zusätzliche Applikationen, Berichte oder automatisierter Auswertungen
- eine Datenexportmöglichkeit (z.B. als CSV oder JSON/XML).
- *keine* Authentifizierungsmechanismen (unbeschränkter Zugang),
- *keine* vorgesehenen Weiterentwicklungen
- *keine* Entwicklung einer Outgoing-API.

Es ist dabei zu berücksichtigen, dass sich diese Kostenschätzungen ausschliesslich auf die Kosten für notwendige IT-Dienstleister beziehen. Bundes- und länderinterne Aufwände (wie z.B. die Koordinationsaufwände der IT- und Datenbankspezialisten der Institutionen, welche die Quelldaten „hosten“) sind hier nicht integriert, da davon ausgegangen wird, dass diese im Rahmen der periodischen Leistungsvereinbarungen abgedeckt werden.

Die Kostenschätzung beruht auf einer Aufwandsschätzung, die mit einem fiktiven Kostensatz von 1'400€ pro Tag hochgerechnet wurde. Dieser Tagessatz entspricht erfahrungsgemäss etwa einem Branchenmittel für Informatikdienstleistungen, kann jedoch je nach gewähltem Anbieter sowohl nach oben als nach unten abweichen.

Tabelle 3: Aufwandsschätzungen für die einmaligen technischen Entwicklungskosten für die Varianten 1 und 3 mit Minimum- und Maximumszenario

Kostenschätzung pro Arbeitspaket:	Arbeitsaufwand in Tagen			
	V1-Min.	V1-Max.	V3-Min.	V3-Max.
Einarbeitung und Verständnis bestehender Anwendungsfälle und Informationsmodelle	15	25	15	25
Datenmodell/-architektur aufbauen (Variante 3) bzw. adaptieren (Variante 1&3)	15	25	20	35
Konzeption und Aufbau der benötigten IT-Infrastruktur	10	15	15	20
Etablierung der automatisierten Datenschnittstellen	20	30	25	40
Entwicklung der Benutzeroberfläche	40	70	50	80
Testing & Deployment	10	20	10	20
Dokumentierung, Supportprozesse, Schulungen	5	15	5	15
Summe	115	200	140	235
+ 15% Projektmanagement und Eventualkosten				

Unter Betrachtung aller technischer Aufwände gehen wir – aufbauend auf das konzipierte Szenario – für die gesamten *einmaligen Entwicklungskosten* für die Variante 1 (Meta-Projektfinder) von etwa 115 bis 200 Arbeitstagen aus. Bei marktüblichen Tagessätzen im Bereich Datenbank- und Webentwicklung (1'400 EUR pro Tag) ist also mit Kosten von rund 185'000 EUR bis 322'000 EUR zu rechnen, wenn man etwa 15% der Projektkosten für Projektmanagement und Eventualkosten, also Kosten für unvorhersehbare Aufwände, addiert. Nicht enthalten in dieser Kostenschätzung sind mittelfristige Anpassungs- und Supportprozesse (Weiterentwicklungen, Bug-fixes, etc.). Dies liegt zum einen daran, dass diese Kosten in erheblichem Masse von der Initialentwicklung abhängen. Je präziser die ursprüngliche Projektdefinition und je hochwertiger die Umsetzung, desto geringer sind die Folgekosten im Bereich Bug-fixing. Zum anderen hängen die Aufwände von sich ggf. verändernden Ansprüchen und späteren Anwendungsfällen ab. Sollten z.B. zu einem späteren Zeitpunkt nicht ursprünglich vorhergesehene Anwendungsfälle gewünscht und von der Trägerinstitution genehmigt werden, können diese hier noch nicht abgeschätzt werden.

Die Aufwände für die *Variante 2* (Ergänzung Transparenzdatenbank) lassen sich aufgrund der speziellen Ausgangssituation der Transparenzdatenbank von aussen nicht plausibel einschätzen. Es ist jedoch davon auszugehen, dass die Kosten aufgrund der hohen Komplexität und des anderweitig ausgerichteten Zwecks der Transparenzdatenbank (vgl. Kapitel 3.2.2) vermutlich wesentlich höher liegen, als dies bei den Varianten 1 und 3 der Fall ist. Dies hängt auch damit zusammen, dass es sich bei dieser Variante

um eine Anpassung eines bereits laufenden Systems handelt. Da bei derartigen Projekten eine Reihe von Abhängigkeiten berücksichtigt werden müssen, sind sie erfahrungsgemäss aufwendiger, als eine Neuentwicklung. Hinzu kommt, dass die Realisierung dieses Szenarios vermutlich nur dann durchgeführt wird, wenn eine *Grundsatzentscheidung* zur Erweiterung der Transparenzdatenbank um weitere Meta-Felder für *andere* Leistungsangebote der öffentlichen Hand getroffen wird (vgl. Kapitel 3.2.2).

Bei Variante 3 ist im Vergleich zu Variante 1 mit Mehrkosten in den Bereichen Datenmodell/-architektur und in der Entwicklung der Benutzeroberfläche zu rechnen. Das hängt vor allem damit zusammen, dass – anders als in Variante 1 – die Daten in dieser Variante effektiv bezogen und gespeichert werden und durch zusätzliche Auswertungsmöglichkeiten und Darstellungsformen des Graphdatenbankssystems ein *erheblicher Mehrwert* entsteht. Die vorbereitenden Arbeiten, sowie das Testing, Deployment und die Supportprozesse entsprechen jedoch grundsätzlich dem Vorgehen von Variante 1 und sind mit vergleichbaren Aufwänden verbunden. Insgesamt schätzen wir für die einmalige Entwicklung der Variante 3 einen Aufwand von etwa 140 bis 235 Arbeitstagen oder Kosten von etwa 225'000 EUR bis 380'000 EUR (bei Berücksichtigung von 15% Projektmanagement- und Eventualkosten).

Bei beiden Kostenschätzungen ist es sehr wichtig zu berücksichtigen, dass erhebliche Unsicherheiten bzgl. der Kosten und des Leistungsumfangs bestehen. Je nach Spezifikation und der Menge der zu berücksichtigenden Institutionen können die Kosten ein Vielfaches der hier angegebenen Schätzungen betragen.

Vergleichende Nutzenbetrachtung

Der Nutzen einer F(F)DB ist je nach Anspruchsgruppe unterschiedlich. Die Einrichtung einer ausschliesslichen FFDB bietet jedoch einen Grundnutzen, der unabhängig davon ist, welcher technische Ansatz gewählt wird. Dieser Grundnutzen liegt darin, dass ein Kerndatensatz Informationen zu den meisten Projekten zugänglich macht, die im Rahmen öffentlicher Programmforschung der an das System angeschlossenen Institutionen finanziert. Dies ermöglicht es den verschiedenen, an diesen Daten interessierten Anspruchsgruppen, ihre Informationsbedürfnisse schnell, vollständig und transparent zu befriedigen.

Betrachtet man die Varianten im Vergleich und nimmt Variante 1 als Referenzszenario, so zeigt sich, dass Variante 2 keinen nennenswerten organisatorischen oder inhaltlichen Mehrwert bringt. Im Gegenteil: um den Nutzen der Varianten 1 oder 3 zu erreichen, wären *erhebliche* Anpassung an der existierenden Datenstruktur, den Authentifizierungsmechanismen, den Abfrage- und Auswertungsfunktionalitäten sowie an den Datenschnittstellen und -quellen notwendig (vgl. Kapitel 5.2). Es ist zudem fraglich, ob es überhaupt sinnvoll möglich ist, die Transparenzdatenbank – deren Zweck grundsätzlich von den Anwendungsfällen einer F(F)DB abweicht – so zu ergänzen und zu verändern, dass beide Zwecke in Einklang gebracht werden können. Variante 2 unterscheidet sich von den Varianten 1 und 3 insbesondere dadurch, dass für die Transparenzdatenbank bereits eine vorhandene Struktur besteht (Infrastruktur, Personal, Rechtsgrundlagen, Datenschnittstellen, etc.). Ob dies jedoch vor dem Hintergrund der unterschiedlichen Ziele der Transparenzdatenbank und einer F(F)DB ein Vorteil oder eher ein Hindernis ist, bedarf einer vorherigen Prüfung.

Die unterschiedlichen Nutzen der Variante 1 und 3 ergeben sich aus dem Umstand, dass in Variante 1 ausschliesslich auf Daten der Quellinstitutionen zugegriffen wird (es werden *Abfragen* übersetzt), während in Variante 3 die Daten effektiv bezogen, selbst verwaltet und abgelegt werden (es werden *Datenstrukturen* angepasst). In Kombination mit der Graphdatenbankstruktur ermöglicht dies bei Variante 3 eine erheblich *perfor-*

*mantere Abfrage komplexer Datensätze*²⁵ als in Variante 1. Zudem ergeben sich angeschlossene Transformations- und Manipulationsoperationen, die von grossem Wert für zusätzliche Auswertungsoptionen oder auf das System aufbauende Applikationen und automatisierte Berichterstattung sein können. Hinzu kommt, dass Graphdatenbanken hocheffiziente Abfragen für komplexe und vernetzte Fragestellungen erlauben. In erster Linie für hochdimensionale Objekte wie Forschungsprojekte, bietet eine Graphdatenbank Auswertungsoptionen, die bisher nur mit erheblichem Aufwand möglich waren. Zudem können Graphdatenbanken einfach und in kleinen Schritten erweitert werden; sie sind sozusagen «schema-frei». Die traditionelleren relationalen Datenbanken (mit Tabellen) brauchen dagegen aufgrund ihrer «schema-pflicht» erheblich rigidere Strukturen, die bei Erweiterungen oder Änderungen häufig aufwändige Anpassungen benötigen, welche teuer, langsam und störend sind. Die Variante 3 bietet damit auf inhaltlicher (Abfragen, Transformationen und Manipulationen) und organisatorischer (Skalierbarkeit, Flexibilität, Umgang mit heterogenen Datenquellen) Ebene einen Mehrwert gegenüber Variante 1. Es ist ausserdem zu beachten, dass die Variante 1 nur bei bereits existierenden Schnittstellen der Quellinstitutionen auf existierenden Abfragemechanismen zugreifen kann. Wenn die Quellinstitution *keine* derartigen Mechanismen hat, ist die Etablierung der Datenschnittstellen in Variante 1 und 3 ähnlich aufwendig.

Der grösste Unterschied zwischen der Variante 4 und einer nationalen Lösung, wie in den Varianten 1-3 vorgesehen, liegt darin, dass sich der *Inhalt* der Datenbank und damit die Nutzungs- und Auswertungsmöglichkeiten der Daten erheblich ausweiten würden, da Daten *aller* angeschlossenen Länder und ihrer jeweiligen nationalen FTI-Systeme vorliegen würden. Zum einen ist dies in der Konzeption aufwendiger, da Mehrsprachigkeit und heterogene Datenquellen berücksichtigt werden müssten, zum anderen würden durch internationale Vergleichbarkeit und Recherchemöglichkeiten für die Anspruchsgruppen ein grosser Mehrwert geschaffen. Da die Variante 4 mit den Varianten 1 bis 3 kombinierbar ist, „erbt“ sie ihre Vor- und Nachteile. Es ist jedoch davon auszugehen, dass die Vor- und Nachteile jeder Variante aufgrund der höheren Komplexität und des höheren Umfangs stärker ausgeprägt wären, als wenn sie ausschliesslich in einem nationalen Rahmen realisiert werden.

6. Internationale Case Studies

6.1 Case Study: Schweiz

6.1.1 Institutionen der Forschungsförderung

In der Schweiz ist der Bund durch die Bundesverfassung verpflichtet, die wissenschaftliche Forschung und Innovation zu fördern. Die Finanzierung der Forschung erfolgt nach liberalen Grundprinzipien; das heisst, Gelder werden aufgrund von Eigeninitiative der Forschenden nach dem Wettbewerbsprinzip vergeben. Ausschlaggebend ist die Qualität der eingereichten Anträge. Ein weiterer Eckpfeiler der Schweizer Forschungspolitik besteht in der Förderung der internationalen Zusammenarbeit.

Der Bund fördert die Forschung über folgende Institutionen und Instrumente:

- Der *Schweizerische Nationalfonds* (SNF, <http://www.snf.ch>) fördert die *freie Grundlagenforschung* und führt *Programmforschung* – Nationale Forschungsschwerpunkte

²⁵ Dies spielt vor allem bei einer Ausweitung der Datentiefe (vgl. Kapitel 4.2) und der Datenbreite um weitere Institutionen eine grosse Rolle.

(NFS) und Nationale Forschungsprogramme (NFP) – in Bereichen durch, wo wirtschaftliche und gesellschaftliche Bedürfnisse gezielt wissenschaftliche Resultate benötigen oder in denen strukturelle Schwächen des Wissenschaftssystems einen konzentrierten Einsatz verlangen.

- Die *Kommission für Technologie und Innovation* (KTI, www.kti.admin.ch) ist zuständig für die *anwendungsorientierte Forschung und Entwicklung* (aF&E), die *Stärkung des Unternehmertums* und den Aufbau von Jungunternehmen.
- Die *Ressortforschung* der einzelnen Bundesämter und die internationale Zusammenarbeit im Rahmen internationaler Forschungsprogramme und internationaler Forschungsorganisationen sowie im Rahmen bilateraler Programme mit Schwerpunktländern sind weitere wichtige Aktivitäten.
- Zudem finanziert der Bund den Verbund der Akademien der Schweiz (www.akademien-schweiz.ch) und unterstützt knapp 30 Forschungsstätten ausserhalb der Hochschulen. Darüber hinaus fördert der Bund die Lehre und Forschung an den Institutionen des ETH-Bereichs (www.ethrat.ch).

Die Kantone engagieren sich für die Forschung vor allem in ihrer Funktion als Träger der Universitäten und Fachhochschulen.

6.1.2 Forschungsförderungs-Datenbanken

In der Schweiz *existiert keine zentrale FFDB*. Die Forschungsprojekte, die vom Bund unterstützt werden, sind in *verschiedenen Informationssystemen* erfasst – je nachdem über welche Institution die Förderung läuft. Zurzeit ist die Errichtung einer übergeordneten Datenbank nicht geplant.

Zu den wichtigsten Forschungsdatenbanken in der Schweiz zählen folgende Einrichtungen:

ARAMIS (www.aramis.admin.ch)

- Das Informationssystem ist seit 1997 in Betrieb. Der Einführung sind mehrere parlamentarische Vorstösse vorangegangen, die eine Verbesserung der Kooperation in der Ressortforschung verlangt haben.
- Die Ziele und Aufgaben sind in einer Verordnung genau beschrieben. ARAMIS will Transparenz schaffen, Doppelspurigkeiten vermeiden und Forschungsprojekte einfach verwalten.
- Die Datenbank enthält Informationen über Forschungsprojekte und Evaluationen, die der Bund selber durchführt oder finanziert. ARAMIS ist webbasiert und besteht aus einem Teil mit öffentlichen Informationen und einem geschlossenen/internen Bereich mit ergänzenden Informationen (Login-Bereich).
- ARAMIS umfasst Informationen zu Projekten der Ressortforschung und Bundesämter sowie jene der KTI. Die einzelnen Bundesorganisationen sind über eine SAP-Schnittstelle angebunden. Sie sind für die Dateneingabe verantwortlich und haben die „Datenherrschaft“. Wenn also Informationen angefragt werden, die nicht über den öffentlichen Teil abgerufen werden können, entscheidet das Bundesamt über den Zugang. Will bspw. ein Journalist wissen, wer die Vertragspartner sind und wie viel Geld an wen fliesst, dann wird der Journalist an die Ämter verwiesen.
- Die Forschungsprojekte des SNF sowie der Hochschulen sind nicht einzeln erfasst. Es ist lediglich die Information zum Gesamtbetrag, der pro Jahr vom Bund an die jeweiligen Institutionen für Forschung fliesst, verfügbar.

P3 (p3.snf.ch)

- Der SNF publiziert alle Forschungsprojekte in der 2012 lancierten Datenbank P3. Sie löste 2012 das bestehende Informationssystem ab. Zurzeit enthält P3 über 66'000 Projekte, an denen rund 95'000 Personen beteiligt sind.
- P3 hat den Auftrag, die vom SNF geförderten Projekte umfassend und transparent darzustellen. Sämtliche in P3 enthaltene Informationen sind öffentlich und können mit anderen Informationen, z.B. von Hochschulen, verknüpft werden. Zum Teil bestehen bei Universitäten und Forschungsinstitutionen Schnittstellen, um P3- Informationen herunterzuladen.
- P3 enthält u.a. Informationen zum Inhalt und auch zum Output der vom SNF unterstützten Projekte. Sämtliche Informationen sind öffentlich. Es besteht keine Schnittstelle zur Datenbank ARAMIS. Der Bundesbeitrag an den Nationalfonds wird als Gesamtbetrag in ARAMIS ausgewiesen.

Weitere Forschungsdatenbanken

- forschungsportal.ch: ist ein gemeinsames Portal von verschiedenen Schweizer Universitäten und verknüpft die Forschungsdaten einzelner kantonaler Universitäten auf einer gemeinsamen Webinterface. Dies ermöglicht es, gleichzeitig in allen Datenbanken nach Forschungsprojekten zu suchen.
- Forschungsdatenbanken der ETH, Universitäten und Fachhochschulen sowie weitere Forschungsorganisationen: die Hochschulen betreiben jeweils eigene Datenbanken, in denen alle Forschungsprojekte der jeweiligen Institution unabhängig von der Förderinstitution erfasst sind.

6.1.3 Daten

Datenstruktur

Da in der Schweiz eine übergeordnete FFDB fehlt, sind die Oberflächen und Datenstrukturen der verschiedenen Informationssysteme nicht harmonisiert. Ähnliche Grundinformationen sind aber in jeder Datenbank zu finden. Dazu gehören insbesondere folgende Angaben:

- Titel
- Laufzeit
- Projektverantwortliche und -beteiligte
- Projektziele
- Abschlussbericht
- Finanzdaten

ARAMIS liegt NABS²⁶ zugrunde. Dieser Standard bezeichnet den Politikbereich, zu dem ein Projekt zu zählen ist. Diese Klassierung ermöglicht es dem Bundesamt für Statistik, internationale Vergleichswerte zu berechnen.

P3 orientiert sich stark an der Open Researcher and Contributor ID (ORCID). Ziel ist es, dass alle Forschenden, die vom SNF gefördert werden, über ORCID eindeutig über Systeme hinweg identifizierbar sind. Der SNF nutzt zudem GRID, eine globale und of-

²⁶ Nomenclature for the Analysis and Comparison of Scientific Programmes and Budgets (Systematik zur Analyse und zum Vergleich der wissenschaftlichen Programme und Haushalte) vgl. <http://www.oecd.org/science/inno/43299905.pdf>

fene Datenbank, die Informationen über Forschungsinstitutionen auf der ganzen Welt enthält. GRID bietet einen Überblick über abgeschlossene und laufenden Forschungsprogramme bzw. -vorhaben und ermöglicht Vergleiche und Auswertungen.

ARAMIS arbeitet auf der Basis der ARAMIS-Verordnung, in der die Ziele, Aufgaben und Zuständigkeiten der Datenbank detailliert beschrieben werden. Die rechtliche Grundlage für P3 bildet das Bundesgesetz über das Öffentlichkeitsprinzip der Verwaltung (Öffentlichkeitsgesetz). Alle Personen erhalten danach grundsätzlich Zugang zu jeder Information und zu jedem Dokument der Bundesverwaltung. Weitere Rahmenbedingungen sind im Bundesgesetz über die Förderung der Forschung und der Innovation (FIFG) festgelegt.

Nutzerinnen und Nutzer

Die FFDB stehen grundsätzlich jeder Bürgerin und jedem Bürger offen. Sowohl ARAMIS als auch P3 führen keine systematischen Nutzer-Auswertungen durch, wissen aber, dass sie einerseits als Informationsquelle für statistische und weiterführende Auswertungen für nationale und internationale Vergleiche sowie für Evaluationen genutzt werden. Andererseits stellen sie wichtige Informationen für Vertreter aus Hochschulen, Politik und Verwaltung bereit, die sich einen Überblick über abgeschlossene und laufende Projekte verschaffen wollen.

Viele Nutzer gelangen über eine Google-Suche zu einem bestimmten Thema direkt auf die beiden Datenbanken.

Vor- und Nachteile

Vorteile

ARAMIS und P3 sind gut etablierte FFDB. Die Akteure sind bekannt. Es gibt einen regen Austausch mit den Forschungsverantwortlichen der Bundesämter (ARAMIS) bzw. den Forschenden (P3), die für die Inhalte des Informationssystems verantwortlich sind. Die Qualität der Daten ist relativ gut. Die Datenbankbetreiber sehen sich in der Funktion von Beratenden, die bei der Eingabe im Falle von Problemen jederzeit zur Verfügung stehen. Die Entscheidungswege sind kurz und die Datenbank-Teams gut eingespielt. Beispielsweise können Systemanpassungen rasch und unkompliziert erfolgen.

Nachteile

Die verschiedenen Forschungsförderungs-Datenbanken sind unabhängige Teile des Gesamtsystems Forschung und funktionieren komplett unabhängig voneinander. Es gibt keine Schnittstellen, der Austausch zwischen den Repräsentanten der verschiedenen Plattformen ist kaum existent. Nationale und internationale Auswertungen und Vergleiche erfolgen meist in aufwändiger Kleinarbeit und mit nicht sehr präzisen Schätzungen.

Fazit

Eine übergreifende Datenbank könnte gemäss dem ARAMIS-Vertreter durchaus Sinn machen. Dafür müssten in einem ersten Schritt die Gemeinsamkeiten der bestehenden Forschungsförderungs-Datenbanken definiert und auf dieser Basis eine Meta-Datenbank konzipiert werden. Wichtig sei, dass eine solche Meta-Datenbank einfach aufgebaut sein müsste, sodass für die verschiedenen Nutzer und Inhaltslieferanten kein Zusatzaufwand entsteht.

Der P3-Verantwortliche würde einer internationalen Forschungsförderungs-Datenbank den Vorzug geben. Erste Schritte in diese Richtung sieht er bei der Datenbank Dimen-

sion (<http://www.uberresearch.com/dimensions-for-funders>) erfüllt, einem privatwirtschaftlichen Anbieter. Sie sammelt öffentlich verfügbare Informationen von über 200 Forschungsförderungs-Organisationen und erlaubt – auch aufgrund intelligenter Suchtechnologien wie dem Machine Learning – Institutionen und Länder vergleichend darzustellen.

6.2 Case Study: Deutschland

6.2.1 Institutionen der Forschungsförderung

Die öffentlich finanzierte Forschung hat in Deutschland einen hohen Stellenwert. Bedeutende Akteure der Finanzierung von Forschung sind Bund und Länder. Durch das föderale System der Bundesrepublik Deutschland erhalten Bund sowie Länder die Möglichkeit der Forschungs- und Innovationsförderung.

Zu den wichtigsten Akteuren auf öffentlicher Seite gehört der Hochschulbereich (Universitäten und Fachhochschulen). Zudem existiert ein breites Spektrum an ausseruniversitärer Forschung, die grösstenteils an gemeinsam von Bund und Ländern geförderten Einrichtungen durchgeführt wird. Zu den wichtigsten Institutionen zählen:

- Die *vier grossen Forschungsorganisationen*: die Max-Planck-Gesellschaft (MPG), die Fraunhofer-Gesellschaft (Fraunhofer), die Hermann von Helmholtz-Gemeinschaft Deutscher Forschungszentren (HGF) und die Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz (Leibniz-Gemeinschaft)
- Die *Deutsche Forschungsgemeinschaft* (DFG, <http://www.dfg.de/>): Die DFG ist die Selbstverwaltungsorganisation der Wissenschaft in Deutschland. Die Kernaufgabe der DFG besteht in der wettbewerblichen Auswahl der besten Forschungsvorhaben von Wissenschaftlerinnen und Wissenschaftlern an Hochschulen und Forschungsinstituten und in deren Finanzierung.

Weitere Fördermittel werden durch Bundesministerien oder Bundeseinrichtungen an Stiftungen, Akademien und die Ressortforschung vergeben.

6.2.2 Forschungsförderungs-Datenbanken

Es existiert keine zentrale FFDB in Deutschland. Die zwei wichtigsten und grössten Datenbanken sind der Förderkatalog des Bundes und die Datenbank GEPRIS der DFG:

Förderkatalog (<http://foerderportal.bund.de/foekat/jsp/StartAction.do>)

- Der Förderkatalog ist eine öffentlich verfügbare Datenbank mit mehr als 110'000 abgeschlossenen und laufenden Vorhaben der Projektförderung des Bundes. Der Datenbestand enthält Informationen folgender Bundesministerien: Bundesministerium für Bildung und Forschung (BMBF); Bundesministerium für Umwelt, Naturschutz, Bau und Reaktorsicherheit (BMUB); Bundesministerium für Wirtschaft und Energie (BMWi); Bundesministerium für Ernährung und Landwirtschaft (BMEL); Bundesministerium für Verkehr und digitale Infrastruktur (BMVI) und Bundesministerium der Justiz und für Verbraucherschutz (BMJV)
- Die geförderten Projekte der vier grossen Forschungsorganisationen (MPG, Fraunhofer, HGF, Leibniz-Gemeinschaft) sind im Förderkatalog enthalten.
- Der Förderkatalog in der jetzigen Form besteht seit 2009. Davor wurde einmal im Jahr eine Print-Ausgabe veröffentlicht und später waren die Informationen zudem auf einer CD erhältlich.

GEPRIS (www.dfg.de/gepris)

- Die DFG verfügt seit 1999 über die Förderdatenbank GEPRIS. Sie dokumentiert inzwischen mehr als 100'000 DFG-geförderte Projekte unter Beteiligung von mehr

als 65'000 Wissenschaftlerinnen und Wissenschaftlern, wovon der überwiegende Teil an deutschen Universitäten beschäftigt ist.

- GEPRIS gibt Auskunft über den Inhalt und das Forschungsziel eines Projektes sowie über die an einem Projekt beteiligten Personen und Forschungsstätten. Seit dem Jahr 2010 veröffentlicht GEPRIS zudem Ergebnisse DFG-geförderter Forschung in Form von Abstracts und ausgewählten Publikationsnachweisen.
- GEPRIS ist öffentlich zugänglich. Es gibt aber eine DFG-interne Version mit einem breiteren Informationsspektrum. Der Datenschutz ist gewährleistet, da jeder DFG-Geförderte über einschlägige Merkblätter, im Rahmen der Antragstellung und zuletzt mit dem Bewilligungsschreiben darüber informiert wird, dass er nach Erhalt einer Bewilligung der Veröffentlichung in GEPRIS widersprechen kann.

Weitere Forschungsdatenbanken

Einzelne Bundesländer verfügen über eigene Forschungsinformationssysteme. Die Trägerschaft kann sehr unterschiedlich aufgebaut sein. Ein Beispiel für eine Datenbank auf Bundesländerebene ist Sachsen-Anhalt²⁷. Auch in Bayern gibt es eine Datenbank, die jedoch über das Bayerische Staatsministerium für Umwelt und Verbraucherschutz läuft²⁸ und entsprechende Projekte enthält. In Niedersachsen²⁹ und Rheinland-Pfalz³⁰ gibt es Datenbanken, die die verschiedenen Forschungseinrichtungen und Hochschulen in den jeweiligen Ländern auflisten bzw. die Projekte dieser zusammentragen.

Auch gewisse Hochschulen besitzen eigene Datenbanken oder sind daran, solche Systeme zu initiieren, wie z. B. die Universität Münster³¹.

6.2.3 Daten

Datenstruktur

Da es in Deutschland kein zentrales Forschungsinformationssystem gibt, variieren die in den verschiedenen Informationssystemen enthaltenen Informationen etwas. Generell sind bei GEPRIS sowie im Förderkatalog folgende Informationen zu Projekten enthalten:

- Titel
- Laufzeit
- Antragssteller bzw. Zuwendungsempfänger

Bei GEPRIS sind zudem Projektbeschreibungen sowie bei abgeschlossenen Projekten Projektergebnisse (Abstracts) und relevante Publikationen zu finden. Im Gegensatz zum Förderkatalog verzichtet GEPRIS jedoch auf Angaben zu Finanzierungsdaten (wie bspw. die Höhe der Fördermittel). Grundsätzlich sind die Informationen in GEPRIS etwas detaillierter und informationsreicher als die des Förderkataloges.

Die DFG verfügte seit ihrer Gründung über ein DFG-internes System der Antragsbearbeitung, mit dem für jeden Antrag Stammdaten zentral erfasst wurden. Im Jahr 2005 wurde dieses System stark ausgebaut und als dezentrales System (ElektrA) etabliert.

²⁷ <https://forschung-sachsen-anhalt.de/project>

²⁸ <http://www.stmuv.bayern.de/themen/forschung/projekte/index.htm>

²⁹ <https://www.forschungsprofile-niedersachsen.de/de/recherche/>

³⁰ <https://www.rlp-forschung.de/>

³¹ <https://www.uni-muenster.de/forschungaz/area/3763>

Die Daten werden zuerst in den verschiedenen Antragstellen erfasst und dann in Elektra übergeführt. Das DFG-System greift nicht auf CERIF (Common European Research Information Format) zurück, ist aber weitgehend CERIF-kompatibel. GEPRIS bildet die Daten so ab, wie sie im Ursprungs-System Elektra erfasst werden.

Auch der Förderkatalog verfügt über ein internes System und greift nicht auf Standards zurück.

Nutzer

GEPRIS sowie der Förderkatalog sind öffentlich und stehen interessierten Bürgerinnen und Bürgern zur Verfügung. Intern wird die Datenbank der DFG zur Experten- resp. Gutachter-Recherche genutzt. Auswertungen der Nutzungsdaten zeigen, dass häufig Personen-Profile abgefragt oder Stichwort-Suchen gemacht werden. Aus Rückmeldungen wird ersichtlich, dass Wissenschaftlerinnen und Wissenschaftler über GEPRIS mögliche Projektpartner recherchieren. Mitarbeitende von Ministerien prüfen, welche Projekte in einzelnen Bundesländern laufen. Forschungsreferenten an Hochschulen beziehen Daten bzgl. der Frage, wie ihre Universität bei der DFG vertreten ist. Pressevertreter suchen via GEPRIS Expertinnen und Experten zu bestimmten Themen und Firmenvertreter sehen nach, ob in GEPRIS innovative Projekte ihres Geschäftsfeldes zu finden sind.

Auch beim Förderkatalog nutzen vor allem Wissenschaftler und Forschungseinrichtungen die Datenbank. Angehörige der Ministerien greifen ebenfalls darauf zu, jedoch verwenden sie oft die interne Version des Förderkatalogs, die Zuwendungsdatenbank des Bundes.

Vor- und Nachteile

Vorteile

GEPRIS zeichnet sich durch das grosse Datenvolumen aus. Etwa ein Drittel aller an deutschen Hochschulen durchgeführter Projekte mit Drittmittelfinanzierung sind durch die DFG gefördert und somit bei GEPRIS erfasst. Insofern gleicht GEPRIS annähernd einem nationalen System.

Der Förderkatalog verfügt insgesamt über ein ähnlich grosses Datenvolumen wie GEPRIS. In ihm werden Projektdaten aus dem Hochschulbereich und den vier grossen Forschungsorganisationen erfasst.

Nachteile

Es gibt keine Schnittstellen oder Vernetzungen der Datenbanken. Nach Ansicht des DFG Vertreters liegt die Zukunft aber nicht in einer zentralen Datenbank, sondern eher bei Systemen, die über standardisierte Daten und die intensive Nutzung sogenannter Identifier (DOI für Publikationen, ORCID für Personen, oder IDs für einzelne Institute und Forschungseinrichtungen) vernetzt werden können.

Fazit

Die DFG hat einige Zweifel hinsichtlich der Einrichtung einer nationalen FFDB geäussert.

Einerseits bemerken sie, dass es in einem föderalen System wie in Deutschland kaum vorstellbar ist, dass sich alle zentralen Stellen darauf einigen, ihre Informationen für ein gemeinsames System aufzubereiten und zur Verfügung zu stellen. Andererseits sehen sie eine grosse Herausforderung in der Qualitätssicherung der Informationen. Insbe-

sondere ist es wichtig, dass nach der Primärbefüllung der Datenbank ein Konzept vorliegt, wie die Informationen laufend zu aktualisieren und in ihrer Qualität zu sichern sind.

Die DFG ist der Ansicht, dass zukünftig eher europäische als nationale FIS eine Perspektive haben. Ein europäisches FIS würde durch die internationale Sichtbarkeit einen signifikanten Mehrwert schaffen – sowohl für Projekte und die daran beteiligten Forscher, als auch für Förderer, die diese Projekte unterstützen. Ein aktuelles Beispiel dafür wäre das Projekt OpenAire (www.openaire.eu).

Beim Förderkatalog könnte man sich Schnittstellen der Datenbanken gut vorstellen, jedoch sehen auch sie eine grosse Herausforderung in der Realisierung einer nationalen Datenbank. Schon heute deckt der Förderkatalog nicht 100% aller in den beteiligten Ministerien bewilligten Zuwendungsfälle ab, sondern jedes Ressort entscheidet eigenverantwortlich darüber, welche Zuwendungsbereiche in den Förderkatalog gestellt werden. Umso komplizierter würde sich die Vernetzung verschiedener Datenbanken gestalten.

6.3 Case Study: Neuseeland

6.3.1 Institutionen der Forschungsförderung

Der grösste Forschungsförderer in Neuseeland ist das Ministerium für Wirtschaft, Innovation und Arbeit (*Ministry of Business, Innovation and Employment* – MBIE). Es alimentiert die wichtigsten Institutionen in der Forschungsförderung: die *Science and Innovation Group* im MBIE, den neuseeländische Rat für Gesundheitsforschung (*Health Research Council of New Zealand* – HRC, www.hrc.govt.nz) und die nationale Akademie der Wissenschaften (*Royal Society of New Zealand* – RSNZ, www.royalsociety.org.nz):

- MBIE: ist vor allem für die Bereitstellung von Mitteln für Forschung und Technologie im öffentlichen Interesse verantwortlich und unterstützt internationale Programme z. B. in den Bereichen Energie, Infrastruktur, Umwelt, Industrie, Dienstleistung etc.
- HRC: ist für die Verwaltung der Gesundheitsforschung zuständig und investiert in eine breite Palette von relevanten Themen. Zudem unterstützt es die Karrieren von Wissenschaftlern in der Forschung und setzt sich mit ethischen und sicherheitsbezogenen Fragenstellungen auseinander.
- RSNZ: verwaltet unter anderem den Marsden Fonds, der exzellente Grundlagenforschung unterstützt.

Neben dem MBIE sind die vom Bildungsministerium alimentierte Tertiary Education Commission (TEC) und Callaghan Innovation die grössten Forschungsförderer. Weitere Ministerien stellen bescheidene Mittel für spezifische Projekte zur Verfügung.

6.3.2 Forschungsförderungs-Datenbanken

Neuseeland hat bisher keine zentrale FFDB. Die grossen Forschungsförderungsorganisationen verfügen über Systeme zur Verwaltung von Forschungsprojekten. Diese Systeme sind aber keine FFDB im eigentlichen Sinne. Sie enthalten Informationen zu den Projekten und sind zudem ein Instrument, um deren Fortschritt und Abschluss zu verfolgen. Die verschiedenen Systeme sind nicht miteinander verbunden und unterschiedlich aufgebaut. Inkonsistenzen sind sogar innerhalb einer Institution vorhanden und

erschweren einen ganzheitlichen Überblick über die staatliche neuseeländische Forschungsförderung.

Die Informationen der Datenverwaltungssysteme sind darüber hinaus nur schwer öffentlich zugänglich. Das MBIE verfügt über ein öffentliches Portal und HRC sowie RSNZ publizieren Förderungsinformationen auf ihren Websites, jedoch ohne Suchfunktion. Aufgrund des „Official Information Act“ müssen Informationen zu staatlich geförderten Projekten für jede Bürgerin und jeden Bürger abrufbar sein. Zurzeit ist dies allerdings nur auf Anfrage möglich.

Aufgrund dieser unbefriedigenden Situation hat Neuseeland beschlossen, eine Nationale FFDB, das National Research Information System (NRIS)³², zu realisieren. Im Planungs- und Umsetzungsprozess federführend war das MBIE, das eng mit allen relevanten Forschungsförderungsorganisationen zusammenarbeitete, um die verschiedenen und vielfältigen Bedürfnisse bestmöglich zu integrieren und das System rasch zu etablieren.

NRIS wird als Online-Portal umgesetzt und soll folgenden Nutzen bringen:

- Transparenz des Forschungs- und Wissenschaftssystems herstellen
- Förderaktivitäten aller relevanten Forschungsförderungsorganisationen abbilden
- Informationen zu Forschungsprogrammen und -projekten öffentlich machen
- Datenqualität verbessern
- Aufwand für die Datensammlung reduzieren
- Forschende sichtbar machen

Eine erste Version des konzeptionellen Modells wird im Herbst 2017 öffentlich verfügbar sein. Die Lancierung von NRIS ist in der zweiten Hälfte vom Jahr 2018 geplant. Ähnliche FFDB sind bereits in Norwegen (www.cristin.no/english) oder Portugal (www.ptcris.pt/en/hub-ptcris-en) etabliert.

6.3.3 Daten

Datenstruktur

NRIS wird die Daten aller relevanten Forschungsförderungsorganisationen aus allen Wissenschafts- und Forschungsbereichen einheitlich aufbereiten. Die jeweiligen Organisationen werden über Schnittstellen an NRIS angeschlossen. Das System soll folgende Informationen enthalten:

- Projektverantwortliche und -beteiligte
- Projektinhalt und -ziele
- Finanzdaten
- Kooperationen
- Output-Daten

NRIS wird alle Informationen öffentlich zugänglich machen. Lediglich hoch sensible Daten, die im System gespeichert sind, werden nur auf Anfrage zur Verfügung gestellt.

In der Planungsphase von NRIS wurde CERIF als Diskussionsgrundlage für die Entwicklung des Datenmodells genutzt. CASRAI wurde für Definitionen und Konzepte ver-

³² Sämtliche Details zum NRIS sind im „Domain Plan“ <http://www.mbie.govt.nz/info-services/science-innovation/research-and-data/pdf-library/research-science-and-innovation-domain-plan.pdf> enthalten.

wendet. Eine wichtige Rolle spielt ORCID bei der Identifikation der Forschenden. NRIS wird die standardisierte ID vollumfänglich einsetzen.

Nutzerinnen und Nutzer

Die Informationssysteme der verschiedenen Forschungsförderungsorganisationen, die hauptsächlich der Verwaltung der Programme und Projekte dienen, werden zurzeit vor allem von internen Regierungsstellen genutzt, allen voran vom MBIE selbst. NRIS adressiert ein viel breiteres Publikum: Politikerinnen und Politiker, Forschungsförderungsorganisationen, Forschende, Wirtschaft und Industrie sowie die breite interessierte Öffentlichkeit werden das System nutzen können. Zudem wird NRIS der Aufbereitung statistischer Daten dienen und internationale Vergleiche Forschungsförderungs- als auch im Wissenschaftsbereich ermöglichen.

Fazit

Der Aufbau von NRIS wird von allen relevanten Forschungsförderungsorganisationen begrüsst und mitgetragen. Der gesamte Prozess ist umfassend geplant und wird entlang eines Masterplans realisiert. Eine grosse Herausforderung stellt die Integration der Daten aus den sehr unterschiedlichen Systemen sowie deren Komplettierung dar. NRIS wird lanciert, sobald die Daten der wichtigsten Organisationen integriert sind und Schritt für Schritt erweitert. Dieses Vorgehen soll jene Institutionen überzeugen, die der Lösung zurzeit noch skeptisch gegenüberstehen.

NRIS bedeutet für das neuseeländische Wissenschaftssystem einen grossen Schritt nach vorne. Erstmals wird es möglich sein, auf Basis eines gesicherten Datenbestands Aussagen zum Fördersystem im Wissenschaftsbereich zu machen. Sämtliche Anspruchsgruppen werden zukünftig in der Lage sein, sich rasch eine Antwort zu ihren spezifischen Fragen zu verschaffen. Darüber hinaus sind internationale Vergleiche und damit die weltweite Positionierung in der Wissenschafts- und Forschungslandschaft möglich.

7. Fazit

Die aktuelle Situation hinsichtlich Erfassung, Speicherung und Verwendung von Forschungsförderungsinformationen in Österreich ist von einer dezentralen, institutionspezifischen Entwicklung geprägt. Das Resultat ist eine Reihe von unterschiedlich erfassten und strukturierten Datenbeständen, die zu unterschiedlichen Zwecken verwendet werden und nur bedingt öffentlich zugänglich sind. Eine institutionsübergreifende Übersicht oder Analyse der Forschungs(förderungs)informationen in Österreich ist nur mit erheblichem Aufwand und unter Einschränkungen möglich. Um diese Situation zu adressieren und einen Überblick über das gesamte System der österreichischen Forschungsförderung zu erhalten, bietet sich die Einrichtung einer Datenbank an, deren Funktion es ist, diese dezentral vorliegenden Daten zentral abrufbar zu machen.

Für die Einrichtung einer derartigen Variante bieten sich aus technischer und organisatorischer Sicht drei Szenarien an:

1. Aufbau eines *Meta-Projektfinders ohne eigene Datenspeicherung*
2. Ausbau der Transparenzdatenbank des Bundesministerium für Finanzen
3. Aufbau einer neuen *Graphdatenbank mit eigener Datenspeicherung*

Variante 1 und 3 sind sich dahingehend ähnlich, dass bei beiden eine gezielt für die Zwecke einer FFDB (vgl. Kapitel 4) zugeschnittene Lösung neu erstellt würde. Sie unterscheiden sich vor allem hinsichtlich der Systemarchitektur und den Auswertungsmöglichkeiten. Variante 1 ermöglicht ausschliesslich zentrale Abfragen, die von den angeschlossenen Teilsystemen (FWF, FFG, AWS, etc.) für jeden der angeschlossenen Datenbestände bearbeitet und ausgegeben werden. Es werden keine Daten im System des Meta-Projektfinders gespeichert, bearbeitet oder ausgewertet. Varianten 2 und 3 hingegen beziehen die Daten der angehängten Teilsysteme in regelmässigen Abständen und speichern diese in der eigenen Architektur ab. Varianten 2 und 3 ermöglichen damit die Bearbeitung der Datenbestände (z.B. durch die Ergänzung weiterer Informationen), zusätzlich sind automatisierte und vertiefte Auswertungsmöglichkeiten der Datenbestände direkt im System möglich. Hinzu kommt, dass die Varianten 2 und 3 wesentlich schneller arbeiten und flexibler auf Anpassungen reagieren kann. Variante 2 unterscheidet sich grundsätzlich von den anderen beiden Varianten. Im Gegensatz zu Variante 1 und 3 würde Variante 2 eine bestehende Infrastruktur nutzen: die Transparenzdatenbank des BMF. Dies brächte zwar gewisse Vorteile (z.B. bereits vorliegende Datenschnittstellen und Rechtsgrundlagen), wäre jedoch auch mit teilweise erheblichen Herausforderungen verbunden. Dies hängt insbesondere damit zusammen, dass der Zweck, die Nutzer und die Daten der Transparenzdatenbank von denen einer Forschungsförderungsdatenbank abweichen. Es ist offen, ob eine derartig umfassende Anpassung der Transparenzdatenbank für einen Teilbereich der *gesamten* öffentlichen Förderangebote möglich ist. Es ist jedoch unwahrscheinlich, dass gegenüber Variante 1 oder 3 weniger Aufwand entstehen würde.

Ein Blick auf die Entwicklung in anderen Ländern (vgl. Kapitel 3.2 und 6) zeigt, dass in den nationalen FTI-Strategien vermehrt auf Ansätze gesetzt wird, die der Variante 3 entsprechen. Welche Institutionen Teil des Systems sind und ob ausschliesslich Forschungsförderungsinformationen oder auch allgemeine Forschungsinformationen einbezogen werden, ist jeweils unterschiedlich. Grundsätzlich scheinen umfassende nationale Systeme gemäss Variante 3 überzeugende Vorteile zu bieten, die bei Revisionen oder Neuentwicklungen nationaler FTI-Strategien ausschlaggebend waren.

8. Anhang

8.1 Liste Der Gesprächspartner (national)

Bundeministerium für Finanzen

Dr. Rudolf Weninger, Robert Weinzettl

Technische Universität Wien

Dipl.-Ing. Sabine Neff, MSc, Dipl.-Ing. Wolfgang Spreicer, Mag. Dr.rer.nat. Elisabeth Schludermann

Bundesministerium für Verkehr, Innovation und Technologie

Dr. Rupert Pichler

Österreichische Forschungsförderungsgesellschaft

Mag. Michael Binder, Mag. Josef Säckl

Bundesministerium für Wissenschaft, Forschung und Wirtschaft

Sektionschefin Mag. Barbara Weitgruber, MinR Mag. Simone Mesner, Dipl.-Ing. Sascha Saxinger

FWF - Fonds zur Förderung der wissenschaftlichen Forschung

Vizepräsidentin Dr. Artemis Vakianis, Dr. Ralph Reimann

AWS Austria Wirtschaftsservice Gesellschaft mbH

Mag. Marlies Baurecht

Amt der Steiermärkischen Landesregierung, Referat Wissenschaft und Forschung

Mag. Michael Teubl

Christian Doppler Forschungsgesellschaft

DI Mag. Brigitte Müller, Claudia Klackl, Ing. Dominik Hauser

Statistik Austria

Mag. Andreas Schiefer

Rat für Technologie und Forschung

Dr. Johannes Gadner

8.2 Liste Der Gesprächspartner (international)

Deutsche Forschungsgemeinschaft (DFG)

Dr. Jürgen Güdler

Informationsmanagement

Kennedyallee 40

53175 Bonn, Deutschland

und

Informationsservice Projektförderung

im Deutschen Zentrum für Luft- und Raumfahrt e.V. (DLR)

Geb. 39 | Linder Höhe

51170 Köln, Deutschland

Bundesministerium für Bildung und Forschung

Christian Görgen

Referat Z22 - Informationstechnik im BMBF

Heinemannstrasse 2

53175 Bonn, Deutschland

Eidgenössisches Departement für Wirtschaft, Bildung und Forschung WBF Staatssekretariat für Bildung, Forschung und Innovation

Thomas Boschung

Leiter ARAMIS

SBFI Nationale Forschung und Innovation

Einsteinstrasse 2

3003 Bern, Schweiz

Schweiz. Nationalfonds

Benjamin Rindlisbacher

Wildhainweg 3

Postfach

3001 Bern, Schweiz

Ministry of Business, Innovation & Employment

Clinton Watson

15 Stout Street, Wellington 6011

New Zealand

Abbildungsverzeichnis

Abbildung 1: Schematische Darstellung der Forschungsförderung in Österreich (2014)	8
Abbildung 2: Basisdatensatz, Kerndatensatz und Minimaldatensatz im Überblick	19
Abbildung 3: Schematische Darstellung des Status Quo der Datengewinnung	23
Abbildung 4 Schematische Darstellung Meta-Projektfinder ohne eigene Datenspeicherung	25
Abbildung 5 Schematische Darstellung zum Ausbau der Transparenzdatenbank	26
Abbildung 6: Schematische Darstellung zum Aufbau einer neuen Graphdatenbank	28
Abbildung 7 Schematische Darstellung einer EU-weit konsolidierten F(F)DB	30

Abkürzungsverzeichnis

ARAMIS: Administration Research Actions Management Information System der Schweizer Eidgenossenschaft

arXiv: E-print Service der Cornell Universität für Mathematik, Physik, Information, Quantitative Biologie, Quantitative Finanzwissenschaften und Statistik

AWS: Austria Wirtschaftsservice GmbH

B_f.dat: Bundesforschungsdatenbank des BMWFW

BA: Bundeskanzleramt

BMWK: Bundesministerium für Bildung, Wissenschaft und Kultur Österreich

BMF: Bundesministerium für Finanzen

BMLFUW: Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft

BMVIT: Bundesministerium für Verkehr, Innovation und Technologie Österreich

BMWFW: Bundesministerium für Wissenschaft, Forschung und Wirtschaft Österreich

CASRAI: Consortia Advancing Standards in Research Administration Information

CDG: Christian Doppler Forschungsgesellschaft

CERIF: Common European Research Information Format

DaFNE: Datenbank für Forschung zur Nachhaltigen Entwicklung des BMLFUW

DB: Datenbank

eCall: Elektronisches Kundenzentrum der FFG

EFRE: Europäischer Fonds für regionale Entwicklung

EU PM: EU-Performance Monitor

EU: Europäische Union

Europe PMC: Life-Science Publikationsarchiv

F&E: Forschung und Entwicklung

FFDB: Forschungsförderungsdatenbank

FFG-G: Österreichische Forschungsförderungsgesellschaft mbH - Errichtungsgesetz

FFG: Österreichische Forschungsförderungsgesellschaft mbH

FIS: Forschungsinformationssysteme

FOG: Forschungsorganisationsgesetz

FP7: europäischen Forschungs- und Innovationsprogramms 2007-2013

FTE-Stiftung: Nationalstiftung für Forschung, Technologie und Entwicklung

FTI: Forschung, Technologie und Innovation

FWF: Fonds zur Förderung der wissenschaftlichen Forschung

GEPRIS: Geförderte Projekte Informationssystem der Deutschen Forschungsgesellschaft

NFIS: Nationales Forschungsinformationssystem

ÖAW: Österreichische Akademie der Wissenschaften

ÖFOS: Österreichische Systematik der Wissenschaftszweige

OpenAIRE: Open Access Infrastructure for Research in Europe

P³: Projects People Publication Datenbank des Schweizerischen Nationalfonds

RH: Rechnungshof Österreich

SQL: Structured Query Language (eine Datenbanksprache)

TDB: Transparenzdatenbank des Bundesministeriums für Finanzen

TDBG: Transparenzdatenbankgesetz

Literatur

- Ackermann Krzemnicki, S., Hägele, B. F. (2016): Die Standardisierung von Forschungsinformationen an Schweizer universitären Hochschulen. Informationen, Analysen und Empfehlungen. Zugriff am 13.07.2017 unter <http://www.performances-recherche.ch/uploads/Bericht%20Standardisierung%20von%20Forschungsinformationen.pdf>
- Bundeskanzleramt (2017): Transparenzdatenbankgesetz 2012, Fassung vom 29.08.2017. URL: <https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=20008050>, Zugriff am 29.08.2017
- Bundesministerium für Finanzen (2016): Entwurf Grobkonzept: Überlegungen zu einer österreichweiten Forschungsförderungsdatenbank auf Basis der Transparenzdatenbank. Internes Dokument.
- Bundesministerium für Wissenschaft, Forschung und Wirtschaft (2016): Bundesforschungsdatenbank – B_f.dat Benutzer-Handbuch. Zugriff am 16.08.2017 unter <https://oravm13.noc-science.at/apex/r/bfdat/108/files/static/v4Y/BFDAT%20Handbuch%202016.pdf>
- Deutsche Forschungsgemeinschaft (2015): Kennzahlen zur öffentlich finanzierten Forschung in Deutschland. Zugriff am 24.07.2017 unter: http://www.dfg.de/download/pdf/dfg_im_profil/zahlen_fakten/foerderatlas/2015/dfg_foerderatlas_2015.pdf
- euroCRIS (2014): Harmonising and Formalising Research Administration Profiles - CASRAI/CERIF. Zugriff am 23.08.2017 unter <http://dspacecris.eurocris.org/handle/11366/204>
- euroCRIS (2015): OpenAIRE Guidelines for CRIS Managers. Zugriff am 22.08.2017 unter https://zenodo.org/record/17065/files/OpenAIRE_Guidelines_for_CRIS_Managers_v.1.0.pdf
- Eurostat (2008): Comparison between NABS 2007 and NABS 1992. Zugriff am 22.08.2017 unter <http://www.oecd.org/science/inno/43299905.pdf>
- European Commission Community Research (n.d.): Guide to Intellectual Property Rules for FP7 projects. Zugriff am 26.08.2017 unter http://ec.europa.eu/research/participants/data/ref/fp7/89593/ipr_en.pdf
- European Commission Directorate-General for Research & Innovation (2017): H2020 Programme – Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020. Zugriff am 26.08.2017 unter http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
- European Research Council (2017): Multi-Beneficiary Model Grant Agreement. Zugriff am 19.08.2017 unter http://ec.europa.eu/research/participants/data/ref/h2020/mga/erc/h2020-mga-erc-multi_en.pdf
- Institut für Forschungsinformation und Qualitätssicherung (iFQ), Fraunhofer-Institut für Angewandte Informationstechnik (FIT), Geschäftsstelle des Wissenschaftsrates (2015): Spezifikation des Kerndatensatz Forschung. Zugriff am 12.05.2017 unter http://www.kerndatensatz-forschung.de/version1/Spezifikation_KDSF_v1.pdf
- Ministry of Business, Innovation and Employment (2016): 2016 Research, Science and Innovation Domain Plan. Zugriff am 22.08.2017 unter <http://www.mbie.govt.nz/info-services/science-innovation/research-and-data/pdf-library/research-science-and-innovation-domain-plan.pdf>
- Österreichische Forschungsgesellschaft (2017): Forschungserfolge – Der Jahresbericht 2016 der österreichischen Forschungsförderungsgesellschaft FFG. Zugriff am 28.07.2017 unter <https://www.ffg.at/presse/digitalisierung-der-ffg-jahresbericht-2016>
- Rechnungshof (RH) (2016): Bericht des Rechnungshofes. Forschungsfinanzierung in Österreich. *Bund 2016/8*. Zugriff am 12. August 2016 unter http://www.rechnungshof.gv.at/fileadmin/downloads/_jahre/2016/berichte/teilberichte/bund/Bund_2016_08/Bund_2016_08_3.pdf
- Task Force FTI (2012): Stellungnahmen der Task Force FTI zu Empfehlungen des Rates für Forschung und Technologieentwicklung. Internes Dokument.
- ÜberResearch (n.d.): Dimensions for Funders. Zugriff am 22.08.2017 unter <https://www.uberresearch.com/dimensions-for-funders/>