



Rat für
I n f o r m a t i o n s
I n f r a s t r u k t u r e n

8

NUTZUNG UND VERWERTUNG VON DATEN IM WISSENSCHAFTLICHEN RAUM

Empfehlungen zur Ausgestaltung von Datendiensten an der
Schnittstelle zwischen Wissenschaft und Wirtschaft

Nutzung und Verwertung von Daten im wissenschaftlichen Raum
*Empfehlungen zur Ausgestaltung von Datendiensten
an der Schnittstelle zwischen Wissenschaft und Wirtschaft*

ABKÜRZUNGSVERZEICHNIS

BBAW	Berlin-Brandenburgische Akademie der Wissenschaften
BMBF	Bundesministerium für Bildung und Forschung
DAI	Deutsches Archäologisches Institut
DFG	Deutsche Forschungsgemeinschaft
DGA	Data Governance Act
DLR	Deutsches Zentrum für Luft- und Raumfahrt
DNG	Datennutzungsgesetz
DPC	Data Publishing Charge
EOSC	European Open Science Cloud
ESFRI	European Stakeholder Forum on Research Infrastructures
FAIR	Findable, Accessible, Interoperable, Reusable
FDZ	Forschungsdatenzentrum
FIZ	Fachinformationszentrum
KI	Künstliche Intelligenz
KMU	Kleine und mittlere Unternehmen
NFDI	Nationale Forschungsdateninfrastruktur
PSI	Private Sector Information
RatSWD	Rat für Sozial- und Wirtschaftsdaten
RfII	Rat für Informationsinfrastrukturen
UrhG	Urheberrechtsgesetz
VG	Verwertungsgesellschaft
VRE	Virtual Research Environments
WR	Wissenschaftsrat (nur bei Literaturangaben)

INHALT

Kurzfassung	1
Executive Summary	3
1. Einführung	5
1.1 Zum gesellschaftlichen und politischen Rahmen	6
1.2 Zum Verständnis von Nutzung und Verwertung im Kontext von Forschungsdaten	8
1.3 Zu Zielstellung und Struktur des Positionspapiers	16
2. Entwicklung in unterschiedlichen Domänen und im Bereich generischer Datendienste	20
2.1 Erd- und Umweltwissenschaften	21
2.2 Naturwissenschaften	24
2.3 Lebenswissenschaften	27
2.4 Sozial-, Verhaltens- und Wirtschaftswissenschaften	30
2.5 Geistes- und Kulturwissenschaften	33
2.6 Long-Tail-Daten aus der Forschung	37
3. Übergreifende Merkmale und Ausprägungen von Datendiensten	40
3.1 Umfang, Technologien und Inhalte der Dienste	41
3.2 Akteure und ihre Rollen im Infrastrukturkontext	44
3.2.1 Datensubjekte – Datenerzeuger – Datennachfrager	44
3.2.2 Geldgeber – Anbieter – technische Dienstleister – Vermittler	47
3.2.3 Qualitätssicherung als Handlungsfeld: die Rolle des Kurators	50
3.3 Governance – zwischen Marktplatz- und Redaktionsmodell	51
3.3.1 Das Forum- oder Marktplatzmodell	52
3.3.2 Das Redaktions- oder Editionsmodell	53
3.3.3 Einbindung der Nutzer in die Governance	55
3.4 Zugangsregimes	56
3.4.1 Zugriff nach Registrierung und auf Antrag	57
3.4.2 Offener Zugang zu Inhalten	58
3.5 Finanzierung von Diensten, Monetarisierung von Daten sowie datenbezogenen Leistungen	59
3.5.1 Zuwendungen aus öffentlicher Finanzierung und Community-Support	60
3.5.2 Monetarisierungsmodelle	62

4.	Zusammenfassung und Bewertung der Befunde	65
4.1	Zusammenfassung	65
4.2	Bewertung	67
5.	Empfehlungen	71
5.1	Selbstbestimmtes Teilen von Forschungsdaten ermöglichen und innerwissenschaftliche Freiräume schützen	74
5.2	Hürden für den Marktzugang kleiner Anbieter und innovativer Dienste senken	76
5.3	Datentreuhandmodelle wissenschaftsfreundlich ausgestalten und fördern – Verfahren der Qualitätssicherung aus der Wissenschaft maßvoll adaptieren	77
5.4	Wissenschaftlichen Zugriff auch bei Monetarisierung von Daten und datenbezogenen Leistungen sichern	79
5.5	Datenverluste vermeiden – dauerhaften Zugang zu Daten und Code vertraglich sichern	79
5.6	Wissenschaftliche Qualitätssicherung gewährleisten – auch auf Seiten des Kooperationspartners	80
5.7	Selektions- und Bewertungskriterien für Forschungsdaten anhand wissenschaftlicher Vorgaben entwickeln	81
5.8	Fachlich-inhaltliche Aufbereitung der Daten für außerwissenschaftliche Kontexte leisten	82
5.9	Beratung und Vermittlung für Fachgemeinschaften organisieren – eine Chance für die NFDI und ihre Konsortien	83
5.10	Essentielle Datenbestände der wissenschaftlichen Fachgemeinschaften und Domänen definieren und sichern	84
6.	Literaturverzeichnis	86
7.	Anhang	89
	A. Alphabetisches Verzeichnis der Fallbeispiele	A-1
	B. Diskussionsimpuls: Datendienste nachhaltig gestalten	B-1
	C. Mitwirkende	C-1

KURZFASSUNG

Datendienste haben für die Forschung wichtige Funktionen. Bei fortschreitender Digitalisierung der wissenschaftlichen Arbeit werden sie zu unverzichtbaren Grundvoraussetzungen für den Erkenntnisgewinn. Sie ermöglichen Zugang zu Daten, verknüpfen, verarbeiten (z.B. analysieren) und archivieren Daten. Dabei können Datendienste unterschiedlichen Zielsetzungen und Zwecken dienen beziehungsweise für unterschiedliche Zielgruppen (wissenschaftliche Fachgemeinschaften, Domänen aber auch öffentliche und private Unternehmen) aufgebaut sein und verwendet werden.

Der Rfll unterscheidet Zwecke der Nutzung, der Verwertung und der Vermarktung von Daten, die Einfluss auf Strukturen, Zugänge und letztlich auf Betriebs- und Geschäftsmodelle der Dienste haben. Je nach Ausgestaltung der Dienste, ihrer Geschäftsmodelle und ihrer Angebote können sich aus derartigen Infrastrukturen für die Wissenschaft unterschiedliche Folgewirkungen ergeben. Der Zusammenhang der konkreten Ausgestaltung einschließlich des „Wirtschaftens“ von Datendiensten mit der Entwicklung digitaler Forschung wird im gegenwärtigen Forschungshandeln noch kaum reflektiert.

Um an dieser Stelle mehr Licht ins Dunkel zu bringen, hat sich der Rfll gut vierzig Datendienste an den Schnittstellen zwischen Wissenschaft, Wirtschaft, Zivilgesellschaft und Verwaltung angesehen. Exemplarisch kamen dabei sechs breitgefächerte Domänen in den Blick: die Erd- und Umweltwissenschaften, die Naturwissenschaften, die Lebenswissenschaften, die Sozial-, Verhaltens- und Wirtschaftswissenschaften, die Geistes- und Kulturwissenschaften sowie heterogene Datenspeicher, die mit den sogenannten „Long-Tail-Daten“ umgehen.

Die in den Domänen ausgewählten Dienste wurden als Fallbeispiele – ohne Anspruch auf Repräsentativität – dahingehend analysiert, ob ihre jeweiligen Betriebs- und Geschäftsmodelle Auswirkungen auf das wissenschaftliche Arbeiten mit „ihren“ Daten haben – z.B. (kosten-)freie Zugänglichkeit oder Archivierungsmöglichkeiten, die dauerhaft einen offenen und qualitätsgesicherten Zugang für Wissenschaftlerinnen und Wissenschaftler ermöglichen.

Im Ergebnis kann festgestellt werden, dass sich die Frage danach, wie wissenschaftskonform oder auch der Forschung zuträglich spezifische Betriebs- oder Geschäftsmodelle von Datendiensten sind, nicht pauschal beantworten lässt – weder für einzelne Domänen, noch für die Wissenschaft insgesamt. Auch auf die Frage, ob im Sinne eines nachhaltigen Datenökosystems für die Wissenschaft Dienste besser öffentlich, besser in privater Trägerschaft (also: kommerziell ausgerichtet) oder besser in Public-Private-Partnership angeboten werden sollten, gibt es keine einfache Antwort.

Für eine angemessene Bewertung kommt es vor allem auf den Nutzungszweck an, mit welchem die Forschung auf einen Dienst zugreift. Ebenso zählt – gerade aus der Sicht auf das gesamte Wissenschaftssystem und seine Datenumwelt – die Relevanz und konkrete Beurteilung von Nachhaltigkeitsdimensionen. In der Summe ist das Prinzip der besten Eignung zielführend. Mischformen, in denen öffentlich geförderte Angebote mit kommerziellen Angeboten verzahnt werden können, haben nach Einschätzung des RfII ein großes Potential für die Wissenschaft. Die Partizipation von Fachgemeinschaften und öffentliche Aufmerksamkeit können den Bestand solcher Arrangements auch „politisch“ sichern. Zentral für ein Gelingen der Auswahl nach bester Eignung bleibt eine große Verantwortung der öffentlichen Hand. Sie muss ein dauerhaftes und nachhaltiges Basisangebot für die Wissenschaft, das mindestens die Funktionalitäten des Suchens und Findens sowie des Bewahrens (Langzeitspeicherung auf maschinenlesbaren Medien) abdeckt, aktiv sicherstellen.

Basierend auf dieser Bewertung der Befunde, gibt der RfII Empfehlungen

- zur staatlichen Regulierung des Marktes für kommerzielle und öffentlich finanzierte Datendienste,
- zur Gestaltung der Kooperationsbeziehungen zwischen wissenschaftlichen Akteuren und externen Anbietern von Diensten und
- zur (Selbst-)Organisation von Handlungsfähigkeit auf Seiten der Wissenschaft.

Mit seinen Empfehlungen steckt der RfII einen Rahmen ab, welcher Wettbewerb und Pluralität unter den Datendiensten und ihrer jeweiligen Geschäftsmodelle zum größtmöglichen Nutzen der Wissenschaft gewährleistet. Einseitige Abhängigkeiten durch Lock-Ins und Monopolstellungen von Anbietern oder Angeboten gehen zu Lasten der wissenschaftlichen Souveränität. Das gilt es auszuschließen. Nicht zuletzt sieht der RfII die Nationale Forschungsdateninfrastruktur (NFDI) und ihre Konsortien im Bereich der Nutzung und Verwertung von Forschungsdaten in einer Schlüsselposition, wenn es im Bereich der Datendienste um die Vermittlung zwischen der Rationalität der wissenschaftlichen und der ökonomischen Sphäre geht.

EXECUTIVE SUMMARY

Data services have important functions for research. As the digitisation of scientific work progresses, they are becoming indispensable basic requirements for gaining scientific knowledge. They provide access to data, link, process (e.g., analyse) and archive data. Data services can serve different objectives and purposes and can be designed and used for different target groups (scientific communities, domains, but also public and private companies).

The RfII distinguishes between purposes of use, exploitation and marketing of data, which have an influence on structures, access and ultimately on operating and business models of the services. Depending on the design of the services, their business models and their offerings, these infrastructures can have different consequences for science. The connection between the concrete design, including the business activities of data services, with the development of digital research is hardly ever reflected in current research practice.

In order to shed more light on this issue, the RfII looked at more than forty data services at the interfaces between science, business, civil society and administration. In this way six wide-ranging domains came into view: the earth and environmental sciences, the natural sciences, the life sciences, the social, behavioural and economic sciences, the arts and humanities, as well as heterogeneous data repositories that deal with so-called „long-tail data“.

The services selected in the domains have been analysed as case studies – without claiming to be representative – to determine whether their respective operating and business models have an impact on scientific work with „their“ data – e.g. (cost-)free accessibility or archiving options that permanently enable open and quality-assured access for scientists.

As a result, it can be stated that the question as to how science-conformant or even research-beneficial specific operating or business models of data services are cannot be answered in a blanket manner – neither for individual domains, nor for science as a whole. There is also no simple answer to the question of whether, in the interests of a sustainable data ecosystem for science, services should be better offered publicly, better in private ownership (i.e.: commercially oriented), or better in public-private partnership.

For an appropriate assessment, it is above all the purpose of use with which the research accesses a service that counts. Equally important – especially from the perspective of the entire scientific system and its data environment – is the relevance and concrete assessment of sustainability dimensions. In sum, the principle of best fit is most purposeful. Mixed forms, in which publicly funded offerings can be interlocked with commercial offerings, have great potential for

science in the estimation of the RfII. The participation of professional communities and public attention can also „politically“ secure the continued existence of such arrangements. Central to the success of selection by best fit remains a major responsibility of the public sector. Public authorities must actively ensure a permanent and sustainable basic service for science that covers at least the functionalities of searching and finding as well as preservation (long-term storage on machine-readable media).

Based on this assessment of the findings, the RfII provides recommendations

- on government regulation of the market for commercial and publicly funded data services,
- on the design of cooperative relationships between scientific actors and external service providers, and
- on the (self-)organisation of agency on the part of the scientific community.

With its recommendations, the RfII defines a framework that ensures competition and plurality among the data services and their respective business models for the greatest possible benefit of science. One-sided dependencies through lock-ins and monopoly positions of service providers or offers are to the detriment of scientific sovereignty. This must be ruled out. Last but not least, the RfII sees the National Research Data Infrastructure (NFDI) and its consortia in the area of the use and exploitation of research data in a key position when it comes to mediating between the rationality of the scientific and the economic spheres in the area of data services.

1 EINFÜHRUNG

Wissenschaftliches Arbeiten bedient sich heute einer großen Zahl an Datendiensten, die ihren Ursprung in ganz unterschiedlichen gesellschaftlichen Sektoren haben, die mit teils vielfältigen Geschäftsmodellen betrieben werden und für eine Vielzahl von Zwecken bereitgestellt werden. Nicht immer sind Wissenschaftlerinnen und Wissenschaftler die einzigen Nutzer solcher Dienste. In jedem Falle leisten wissenschaftlich relevante Datendienste einen Beitrag zur Nutzung und Verwertung von Forschungsdaten. Sie bieten Zugänge zu Daten, die für Forschungszwecke interessant sind, sie bieten teilweise auch Werkzeuge an, mit denen diese Daten bearbeitet, verknüpft und analysiert werden können und sie bieten Speicher- und Archivierungsmöglichkeiten, mit denen Forschungsdaten langfristig und sicher bewahrt und zugänglich gemacht werden können. Für diese Zwecke hat die Wissenschaft selbst diverse – in der Regel disziplinspezifische oder für ein Forschungsfeld maßgebliche Dienste hervorgebracht. Sie bedient sich aber auch von kommerziellen Anbietern bereitgestellter Dienste oder greift auf andere wissenschaftsexterne Dienste zu. Hinter etlichen Angeboten stehen nicht selten sogenannte Datenintermediäre. Diese Funktion übernehmen z.B. Bibliotheken, Rechenzentren, Datenzentren, Sammlungen, Archive und Wissenschaftsverlage – somit Akteure, die häufig aus dem wissenschaftsbezogenen Bereich der Informationsinfrastrukturen stammen.

Das Feld ist für die wissenschaftliche Forschung außerordentlich bedeutend und gleichzeitig unübersichtlich. Der Rat für Informationsinfrastrukturen (RfII) hat seit Ende 2018 wissenschaftsrelevante Datendienste einer Sichtung unterzogen und mittels der Sondierung von „Nutzung und Verwertung“ sowie der Möglichkeit zur Vermarktung der dort generierten und gespeicherten Daten genauer betrachtet.

Der RfII untersuchte gut vierzig Datendienste an den Schnittstellen zwischen Wissenschaft, Wirtschaft, Zivilgesellschaft und Verwaltung. Die Dienste wurden im Sinne von Fallbeispielen auch dahingehend analysiert, ob ihre jeweiligen Geschäftsmodelle Auswirkungen auf das wissenschaftliche Arbeiten mit „ihren“ Daten haben – z.B. (kosten-)freie Zugänglichkeit oder Archivierungsmöglichkeiten, die dauerhaft einen offenen und qualitätsgesicherten Zugang für Wissenschaftlerinnen und Wissenschaftler ermöglichen. Diese Fragestellung erbringt kleinteilige Antworten. Sie ist aber zweifellos von großer wissenschaftspolitischer Bedeutung, nicht nur, weil hier öffentliche Ressourcen verausgabt werden. Es geht auch darum, angesichts etwaiger Verwertungs- und Vermarktungsinteressen von Diensteanbietern die „Souveränität“ der Wissenschaft im Umgang mit

Sichtung von
Fallbeispielen
für wissenschafts-
relevante
Datendienste

den von ihr erzeugten und prozessierten Daten zu sichern oder auszubauen.¹ Dies betrifft selbstverständlich auch den barrierefreien Zugriff von Wissenschaft auf weitere Daten (z.B. in Wirtschaft und Verwaltung), die sie benötigt, um Erkenntnisfortschritte zu erzielen und ihren Beitrag zu Innovationsfähigkeit und der Lösung der großen gesellschaftlichen Herausforderungen zu leisten.

Herangehensweise und Leitfragen

Die Fallbeispiele wurden auf Grundlage öffentlich zugänglicher Dokumente, Webseiten und Informationen recherchiert. Die Herangehensweise sowie Leitfragen wurden von der Arbeitsgruppe „Nutzung und Verwertung“ entwickelt, die der RfII im November 2018 auf seiner 13. Ratssitzung eingesetzt hat. Erste Ergebnisse und daraus resultierende Bewertungen sind bereits in die Erarbeitung der Stellungnahme DATENTREUHANDSTELLEN GESTALTEN – ZU ERFAHRUNGEN DER WISSENSCHAFT (April 2020) sowie den Diskussionsimpuls DATENDIENSTE NACHHALTIG GESTALTEN (Mai 2020) eingeflossen.

1.1 ZUM GESELLSCHAFTLICHEN UND POLITISCHEN RAHMEN

Digitaler Wandel verändert Nutzung und Verwertung von Forschungsdaten

Der digitale Wandel verändert die Gegebenheiten der Nutzung und Verwertung wissenschaftlicher Ergebnisse und der im Forschungsprozess erzeugten Daten. Schon früh wurde mit Digitalität die Erwartung verbunden, Wissen solle weltweit ungehindert zirkulieren können. Offenheit und Transparenz sind hierbei Leitbegriffe geworden. Aber auch die Idee von Wissen als Ressource einer durch Innovationen wachsenden Wirtschaft treibt den digitalen Wandel an – und hier unter dem Stichwort der digitalen und technologischen „Souveränität“ nicht zuletzt auch die Vorstellung, dass Zugriff auf Daten und Regulierungshoheit über Daten und die sie bereitstellenden Dienste Vorteile im internationalen Wettbewerb von Wirtschaftsräumen und politischen Wertesystemen sichern kann.

Open Access und Open Science bedürfen besonderer Standards

Im Zuge der Open Access-Transformation im Wissenschaftssystem sowie der noch weiterreichenden Vision einer „Open Science“ sollen nicht nur herkömmliche Publikationen, sondern auch Forschungsdaten weitgehend offen – sowie idealerweise für die Nutzung kostenfrei – zur Verfügung stehen bzw. zur Verfügung gestellt werden.² Diese Zielsetzung eröffnet grundsätzlich auch kommerziell agierenden Anbietern von Daten und auf Daten aufgebauten Services die Möglichkeit eines Zugriffs auf wissenschaftlich erzeugte Daten. Ein solcher Zugriff kann dann zum Zwecke der wirtschaftlichen Verwendung

¹ Konrad/Förstner et al. (2020) – Digitale Dienste für die Wissenschaft. Hier wird diagnostiziert, dass bei von kommerziellen Anbietern betriebenen Diensten ein Mangel an Datenschutz, Nachhaltigkeit und Schnittstellenkompatibilität zu konstatieren sei. Das Problem entstehender Abhängigkeiten thematisiert RfII (2019) – Herausforderung Datenqualität, S. 68 f.

² Vgl. RfII (2020) – Comment on the UNESCO.

geschehen: etwa durch Transfer an die forschende Industrie bzw. anderweitig tätige Unternehmen oder an weitere Bereiche des öffentlichen Lebens, aber auch durch Aufbereitung, Anreicherung, Speicherung oder Verknüpfung von Daten(-sätzen) mit weiteren Dienstleistungen (z.B. Analyseinstrumenten etc.) als kostenpflichtiges Angebot an die Wissenschaft selbst.

In seinem ersten Positionspapier LEISTUNG AUS VIELFALT (2016) weist der Rfll aus Sicht des Wissenschaftssystems bereits auf Probleme und einen Gestaltungsbedarf für die Offenheit von Forschungsdaten hin. Unter anderem hat der Rfll verbindliche Mindeststandards gefordert, welche die Zugänglichkeit und Nachnutzung publizierter Daten kurz- und mittelfristig sichern.³

Weitergehende Hinweise zur Nutzung und Verwertung von Forschungsdaten sind in einer Rfll-Stellungnahme von 2019 enthalten, die sich auf das Maßnahmenpaket der EU zur Förderung eines einheitlichen europäischen Datenraumes bezieht.⁴ Darin ist insbesondere die Open Data Richtlinie⁵ für das Wissenschaftssystem von hoher Relevanz. In dieser Richtlinie werden erstmals auch Forschungsdaten im Sinne von „public sector information“ interpretiert. Sie sind damit dem Anwendungsbereich der neuen EU-Norm unterstellt worden. Die Open Data Richtlinie regelt zwar zunächst recht allgemein, dass bereits veröffentlichte Forschungsdaten für gewerbliche und nichtgewerbliche Zwecke kostenfrei weiterverwendet werden können, „soweit sie öffentlich finanziert wurden und wenn sie von Forschern, Forschungseinrichtungen oder Forschungsförderungseinrichtungen bereits über ein institutionelles oder thematisches Archiv öffentlich zugänglich gemacht wurden.“⁶ Gleichwohl rückt damit eine über die wissenschaftliche Nutzung hinausgehende Verwendung von Forschungsdaten (analog zu Daten der öffentlichen Verwaltung bzw. der Städte und Kommunen) in den Fokus: Forschungsdaten sollen besser *verwertet* bzw. (durch wirtschaftliche Akteure) *vermarktet* werden können. Die genaue Ausgestaltung von Zugangsregelungen zu den Forschungsdaten hat die EU den Mitgliedstaaten überlassen. Der Kabinettsbeschluss zur Umsetzung der EU-Richtlinie in das deutsche Recht (Datennutzungsgesetz, DNG) folgt weitgehend der EU-Richtlinie und hebt die bisherige Ausnahmeregelung für die Forschung von 2013 auf.⁷

Neue europäische
und nationale
Gesetzgebung
sichert freie Nutzung
von Daten zu

³ Rfll (2016) – Leistung aus Vielfalt, S. 59 f.

⁴ Rfll (2019) – Stellungnahme aktuelle Entwicklungen Open Data.

⁵ Richtlinie des Europäischen Parlaments und des Rates vom 20. Juni 2019 über offene Daten und die Weiterverwendung von Informationen des öffentlichen Sektors (Open Data Richtlinie). EC (2019) – 2019/1024/EU-PSI-Directive.

⁶ Open Data und PSI-Richtlinie, Artikel 10 Abs. 2.

⁷ Eine Lizenzierung bzw. der Erlass von Nutzungsbestimmungen sind zulässig, sie sollen aber objektiv, verhältnismäßig und nichtdiskriminierend sein und dürfen nicht zu einer Wettbewerbsverzerrung führen. Vgl. Bundesregierung (2021) – Entwurf eines Gesetzes, S. 52 f.

Forschungsdaten sind damit auf neuartige und mit Blick auf die Folgen noch kaum absehbare Weise nicht nur zur (wissenschaftlichen) Nutzung, sondern auch zur (wirtschaftlichen) Verwertung bestimmt.

Flankiert wird diese Entwicklung durch weitere europäische Rechtssetzungsakte wie den Data Governance Act (DGA)⁸, den Digital Markets Act (DMA) und den Digital Services Act (DSA), die sich 2021 im europäischen Gesetzgebungsverfahren befinden. Auch sie werden auf die Bereitstellungs- und Nutzungsbedingungen von Forschungsdaten an der Schnittstelle zwischen der Wissenschaft und anderen gesellschaftlichen Sektoren voraussichtlich Einfluss nehmen.

1.2 ZUM VERSTÄNDNIS VON NUTZUNG UND VERWERTUNG IM KONTEXT VON FORSCHUNGSDATEN

Nicht verrechtlichtes
Teilen von Daten im
Wissenschaftssystem

Wissenschaftspolitische Initiativen zu „Open Data“ und „Open Access“⁹ sind wichtige Impulse für die Analyse von Datendiensten unter den Aspekten von Nutzung und Verwertung gewesen. Dennoch werden die Dienste in diesem Positionspapier nicht primär unter dem Gesichtspunkt von „Openness“ betrachtet. Denn schaut man genauer hin, so finden sich im Diskurs über die „Offenheit“ von Forschungsdaten neben der originären Begriffsdefinition, die allein auf Zugänglichkeit abhebt,¹⁰ zahlreiche Konzepte ganz anderer Art: z.B. das „Teilen“ von Forschungsdaten, von der Weitergabe „peer-to-peer“ oder der Nutzung in definierten Teams bzw. Gruppen bis hin zu „public sharing“ (Nutzung zur Veröffentlichung).¹¹ Ebenso wird „offen“ sowohl mit gänzlich kostenlosem als auch mit (z.B. durch Nationallizenzen für Publikationen und Software) für große Nutzergruppen finanziertem Zugang assoziiert.

Freie Datenweitergabe
ist ein Kernbestandteil
des wissenschaftlichen
Wertesystems

„Veröffentlichung“ ist freilich ein Rechtsbegriff, der ursprünglich anderes meint. Das Veröffentlichende von Forschungsdaten durch Wissenschaftler oder wissenschaftliche Einrichtungen ist für sich genommen gerade kein Angebot zur schrankenlosen Verwertung – eher ist das Gegenteil der Fall: Mit der Veröffentlichung wird wissenschaftlich geprüftes Wissen in erster Linie einem Publikum überlassen, das primär selbst aus Wissenschaftlern besteht und Wissen seinerseits zu Forschungs- oder auch Bildungszwecken (nach)nutzt. Intermediäre (etwa Wissenschaftsverlage) können zwar prinzipiell daran verdienen. Auch an einer industriellen Verwertung (bis hin zur Vermarktung) können sich

⁸ Hierzu hat der RfII Stellung genommen, siehe RfII (2021) – Stellungnahme zum DGA.

⁹ RfII (2016) – Leistung aus Vielfalt, Anhang A. Begriffsklärungen, S. A19.

¹⁰ „Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)“ The Open Definition.

¹¹ Vgl. Knowledge Exchange (2014) – Sowing the Seed, S. 22; Whyte/Pryor (2011) – Open Science in Practice, S. 207.

Forschende beteiligen (Patente sichern entsprechende Rechte). Die Wissenschaftlerinnen und Wissenschaftler selbst zielen aber mit der Weitergabe von Wissen (Forschungsdaten und -ergebnissen) zuallererst Reputationsgewinne an – sie akkumulieren, mit dem Soziologen Pierre Bourdieu gesprochen, symbolisches Kapital auf ihrem Feld.

Wissenschaftsintern verwendet man auf dieser Basis traditionell gemeinsame Ressourcen. Zu wissenschaftlichen Zwecken vorgehaltene Archive, Repositorien, Bibliotheken und für ganze Forschungsgemeinschaften/Communities zugängliche Großgeräte belegen das. Die digitale Welt hat das in den ersten Jahren der Digitalisierung (zumindest vordergründig) abgebildet: Fachnahe digitale Repositorien und der Forschung zuarbeitende Datenzentren wollen und sollen niedrigschwellig die Nutzung in der Wissenschaft intensivieren. Sie erschließen aber nicht (primär) an kommerziellen Gewinnchancen orientiertes Wissen für Märkte. Differenzierte Nutzungsbedingungen können mit der Zurverfügungstellung von digitalen (wie zuvor auch nichtdigitalen) Forschungsdaten gleichwohl verbunden sein, von der „Offenheit“ bis hin zu moderaten Gebühren oder auch Auflagen, was die Kompetenzen von Nutzenden oder die Relevanz des Nutzungsanlasses angeht. Es gibt Auffindbarkeits- und Zugangsprobleme, die mit etablierten Subskriptionssystemen verbunden sind, und ebenso kann eine Nutzung digitaler Daten technisch voraussetzungsreich sein. Die seit einigen Jahren zur Qualitätssicherung im Forschungsbereich propagierten FAIR-Prinzipien setzen mit der Optimierung der Findbarkeit, Zugänglichkeit, Interoperabilität und Nachnutzbarkeit an diesem Punkt an. Daten müssen, um FAIR-Bedingungen zu genügen, technisch hindernisfrei genutzt werden können, aber nicht notwendigerweise „offen“ sein. Die FAIR-Prinzipien erlauben vielmehr sehr verschiedene Zugangsregimes.

Niedrigschwellige
Nutzung durch fach-
nahe Intermediäre

Neuere Konzepte von Datentreuhänderschaft, die in der Wirtschaft und im Verbraucherschutz aktuell diskutiert werden, lassen sich demgegenüber sowohl als Regulierungsmechanismen des Datenteilens als auch als Regulierung (oder zumindest: Vorbereitung) einer Daten*verwertung* beschreiben.¹² Sie finden unter nicht-offenen Bedingungen statt und Fragen des Datenzugangs können hier durch marktübliche Instrumente gesteuert werden (Preisbildung, Marketing, Gewinnchancen als Anreiz). Damit unterscheiden sie sich von wissenschaftlichen nicht-offenen Regulierungsinstitutionen wie z.B. Forschungsdatenzentren. Dort stehen neben Aufgaben des Aufbereitens und Archivierens für künftige Fragestellungen sowie des Schutzes sensibler Daten (insbesondere der „Datensubjekte“, siehe. 3.2) auch Motive der Rechenschaftslegung und Überprüfbarkeit (für Ergebnisse) im Vordergrund.

Regulierungsmechanis-
men für das Datenteilen

¹² RfII (2020) – Stellungnahme Datentreuhandstellen sowie Verbraucherzentrale Bundesverband e.V. (2021) – Stellungnahme Data Governance Act.

Von Interesse ist daher die Frage nach methodengerechter Nutzung und Verwertung von Forschungsdaten an der Schnittstelle zwischen Wissenschaft, Wirtschaft, öffentlicher Verwaltung und (Zivil-)Gesellschaft. Um die jeweilige Sachlage beschreiben zu können, wird nachfolgend mit einem Modell der „Offenheit in Schichten“ gearbeitet, das eine Differenzierung verschiedener Teilschritte erlaubt (vgl. Abbildung 1).

Kein Eigentum an,
aber Entscheidungsbefugnisse über die Daten

Grundlegend ist dabei die Annahme, dass es nicht umstandslos ein „Eigentum an Daten“ gibt. Im Forschungskontext werden allerdings gewisse Verfügungsrechte bzw. Entscheidungsbefugnisse über Daten ausgeübt – auch wenn hierzu feste Regeln fehlen und ganz unterschiedliche rechtliche Rahmenwerke dabei aktiviert werden können.¹³ Wissenschaftspolitisch ist von Interesse: Wie über Forschungsdaten verfügt wird, folgt innerhalb des Bezugssystems Wissenschaft einer Logik, die auf das Streben nach Erkenntnisgewinn wie auch nach wissenschaftlicher Reputation ausgerichtet ist. Wo dies nicht der Fall ist, können Störungen für den im wissenschaftlichen Sinne „freien“ (nämlich: gerade nicht kommerziell gelenkten) Austausch auftreten. Das Wissenschaftssystem muss deshalb von anderen Bezugssystemen unterschieden werden – namentlich von der Wirtschaft, deren Funktionsweise im Wesentlichen durch die gewinnorientierte Vermarktung von Produkten gelenkt wird. Wo Wissenschaft und Wirtschaft interagieren um Innovation zu erzeugen, durchdringen sich in der Regel Prozesse der „Verwertung“ (so die Wissenschaftssicht) und solche der „Vermarktung“ (so die Wirtschaftssicht) (Abbildung 1 links und rechts).

¹³ So können Forschungsdaten urheberrechtlich wie auch patentrechtlich oder auch als Objekt professioneller Verantwortung betrachtet werden. Einen Überblick der Rechtslage für Forschungsdaten gibt Lauber-Rönsberg et al. (2018). Forschungsdaten können einer forschenden Person zugeordnet werden. Eine solche Zuordnung befugt den Forscher beispielsweise die Daten zu veröffentlichen oder die Bedingungen der Nachnutzung wirksam zu bestimmen. Diese „Entscheidungsbefugnis“ kann durch Urheberrecht (längere Texte, Zeichnungen, Datenbanken), Patentrecht oder einen Dienstvertrag begründet sein. Auch die „gute wissenschaftliche Praxis“ gilt als mögliche Grundlage der Zuordnung von Daten zum Verantwortungsbereich einer Person. Allerdings besteht keine hinreichende Rechtssicherheit bezüglich der Rechte konkreter Individuen oder Institutionen (z.B. dem Recht zur Nutzungsrechtseinräumung). Die Frage der Entscheidungsbefugnisse bezogen auf Primärdaten ist letztlich rechtlich ungeklärt (Lauber-Rönsberg et al. (2018) – Rechtliche Rahmenbedingungen FDM, S. 5). Bei so genannten digitalen Objekten ist die Gestaltung von Kontrollrechten eine Herausforderung. So können Dritte die Kontrolle über ursprünglich von Forscherinnen und Forschern erzeugte Daten übernehmen, z.B. durch Aggregation bibliographischer Daten, die dann in Form kommerzieller Produkte von wissenschaftlichen Bibliotheken zurückgekauft werden (Neylon/Belsø et al. (2019) – Open Scholarship).

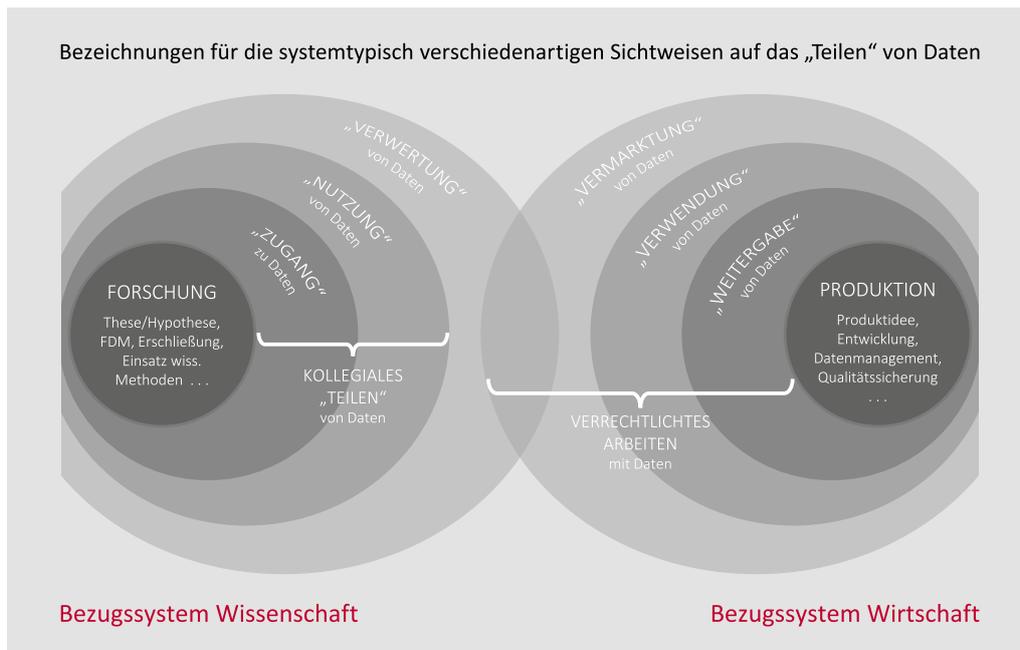


Abbildung 1: Systemtheoretisch inspiriertes Schichtenmodell der basalen Praktiken in der Wissenschaft und in der Wirtschaft (jeweils auf dem Weg zu einer „Offenheit“ von Daten).

Vor dem Hintergrund dieses Schichtenmodells bieten sich folgende begriffliche Unterscheidungen an, die die Kernlogiken (und damit auch die Motivationen und Erwartungshorizonte der dort handelnden Individuen) des Wissenschafts- von dem des Wirtschaftssystems unterscheiden:¹⁴

Handlungs-orientierungen im Wissenschafts- und im Wirtschaftssystem

- Teil des wissenschaftlichen **Forschungsprozesses** ist häufig die methodische Gewinnung, Erschließung und Qualitätssicherung von Daten. Der Begriff „Gewinnung“ umfasst z.B. Messung, Erhebung, Auslesen, synthetische Produktion und ggf. auch Kopieren von Daten (zwecks eigener Nutzung/Forschung). Zum Gestaltungsbedarf bzw. zur wissenschaftlichen Dokumentation und Qualitätssicherung so gewonnener Forschungsdaten hat sich der RfII 2019 ausführlich geäußert.¹⁵
- Der wissenschaftstypische **Zugang** zu Daten (= erste Schicht) umfasst das Finden und den gelingenden Zugriff zwecks Nutzung durch die Forschung.

¹⁴ Wobei zu beachten ist, dass beide Logiken in Innovationsprozessen, in denen Wissenschaft und Wirtschaft strukturell interagieren, miteinander verwoben sind (am dichtesten in der Industrieforschung). Es wird also nicht gegeneinander, sondern miteinander gehandelt – wobei beide Seiten primär ihre Kerninteressen im Blick behalten. Wieweit die wissenschaftliche Eigenlogik hierbei gesellschaftlich oder ökonomisch dienstbar gemacht werden kann und ob die Unterscheidung zwischen Grundlagenforschung („reine“ wissenschaftliche Logik) und anwendungsbezogener Forschung überhaupt noch trägt, ist Gegenstand aktueller wissenschaftssoziologischer und auch wissenschaftspolitischer Debatten (einige Stichworte sind hier: „transformative Wissenschaft“, „Mode 2-Wissenschaft“, „Triple Helix“-Innovationsmodell etc.).

¹⁵ RfII (2019) – Herausforderung Datenqualität.

Zugang setzt Bereitstellung plus Zugänglichmachung (z.B. eine maschinelle oder von Personen bedienbare Schnittstelle) voraus.

- Die wissenschaftliche **Nutzung** („Teilen“) von Daten (= zweite Schicht) umfasst die Verarbeitung zu Forschungszwecken bzw. im Rahmen von Forschungsprozessen. Hierzu zählen Schritte wie Selektion, Transformation, Analyse von Daten. Ziele der Nutzung sind hierbei innerwissenschaftliche: Wissenszuwachs für künftige Forschung („Erkenntnisfortschritt“) sowie mögliche Reputationsgewinne der beteiligten Forscherinnen und Forscher. Wissenschaftsinterne Nutzung von Daten setzt lediglich Autorisierung voraus, aber keine Übertragung weitergehender Verfügungsrechte. Auch fließt kein Geld, es gibt keine wettbewerbliche Preisbildung, es wird nichts verkauft.
- Erst die **Verwertung** (= dritte oder „äußere“ Schicht) umfasst die Weitergabe von Forschungsdaten und -ergebnissen an eine Öffentlichkeit, an die Wirtschaft oder weitere Bereiche der Gesellschaft (z.B. Kunst/Kultureinrichtungen, öffentliches Gesundheitssystem, Behörden und Politik). Sie ist der Beginn eines Innovationszyklus, in dem Daten aus der Wissenschaft in gesellschaftlich weiterverwertbare Informationen – in der Wirtschaft: in voraussichtlich kommerziell erfolgreiche Produkte – transformiert werden. Auf dieser Stufe werden Forschungsergebnisse und -daten (markt)rechtlich gefasst, deshalb kommen hier verschiedene Formen formalisierter Vereinbarungen ins Spiel (Lizenzen, Verträge, Vereinbarungen etc.).

Leitunterscheidung:
Wissenschaftsinterne
Nutzung –
wissenschaftsexterne
Verwertung –
kommerzielle
Vermarktung

Aus diesen Unterscheidungen folgt: Wo eine „Nachnutzung“ von Forschungsdaten in der Forschung erfolgt, geht es erst einmal um die gänzlich wissenschaftsinternen Schichten Zugang und Nutzung, nicht aber bereits um „Verwertung“. Auch das Teilen oder die Veröffentlichung von Forschungsergebnissen wie auch Daten sind typischerweise erst dann „Verwertung“, wenn kommerzielle Partner, die ihrerseits „vermarkten“ (Verlage, Plattformanbieter, *Social Media* etc.), mit beteiligt sind. Ein Beispiel für einen Partner der Wissenschaft, der „Verwertung“ unterstützt, ohne zu „vermarkten“, sind die wissenschaftsinternen Publikationsserver (auch für Daten), z.B. in wissenschaftlichen Bibliotheken. „Verwertung“ wiederum kann (muss aber nicht zwingend) auf einen wirtschaftlich/kommerziellen Zweck ausgerichtet sein. Es geht um „Wertschöpfung“ ganz allgemein. Auch der Einsatz von Forschungsergebnissen in Behörden – beispielsweise zur Kriminalitätsprävention – wäre eine Form von „Verwertung“: nicht „wissenschaftlich“, aber auch nicht kommerziell.

Wirtschaftliche Abnehmer (oder „Nachfragende“) von Forschungsdaten, Forschungsergebnissen und -produkten bewegen sich (und verwenden Daten) in einem anderen Bezugssystem (Abbildung 1 rechts).

- Primär geht es hier um **Produktentwicklung** und die Wertschöpfung aus der Produktion materieller und immaterieller Güter. Hierfür wird zunehmend die eigene Gewinnung von Daten wie auch der Zukauf oder die Übernahme von Daten aus verschiedenen Sektoren inkl. der Wissenschaft notwendig.
- Es kann eine **Weitergabe** erfolgen: Gemeint ist die Überlassung von Daten ohne kommerzielle Verwendung (z.B. kostenlose Abgabe an die Wissenschaft). Wichtig ist hierbei: Schon die Weitergabe, ebenso wie die Verwendung, setzen in diesem Bezugssystem Urheberrecht, Patentrecht etc., mindestens aber belastbare Verfügungsrechte voraus. Die Datenweitergabe aus der Wirtschaft unterscheidet sich somit von dem in der Wissenschaft verbreiteten kollegialen „Teilen“ von Daten.¹⁶
- Eine ökonomische oder **gewerbliche Nutzung („Verwendung“)** umfasst die wirtschaftstypische Verarbeitung von Daten zu Datenprodukten bzw. zu ökonomisch profitablen Gütern oder Diensten. Auch hier müssen Verwendungs- und Verwertungsbefugnisse eindeutig geklärt sein.
- Eine **Vermarktung** von Daten umfasst den Verkauf oder andere Formen des rentablen Einsatzes von Daten oder Datenprodukten/Diensten zwecks Werbung/Kundenbindung oder Gewinnerzielung.

Die heutige internetbasierte Digitalwirtschaft, die Wertschöpfung aus der Verarbeitung und Weitergabe von Daten erzielt, zeichnet sich freilich dadurch aus, dass Prozesse der (kostenfreien) Verwendung und (profitorientierten) Vermarktung von Daten Hand in Hand gehen. Gewinnerzielungsabsichten können dabei in die Zukunft verschoben werden, solange es entsprechenden Unternehmen gelingt, Investoren davon zu überzeugen, dass die Verfügung über Daten und die Häufigkeit ihrer Nutzung bereits für sich betrachtet einen Geldwert darstellen und damit künftigen Mehrwert versprechen.¹⁷ Der „Wert“ dieser Unternehmen und ihrer Wertschöpfung bemisst sich somit anhand der Quantität an Daten und Zugriffen, nicht anhand der damit aktuell erzielten Gewinne. Geschäftsmodelle zielen in diesem Bereich darauf ab, über Daten und mit Daten verknüpfte Services (Produkte) möglichst viele Nutzer zu binden und – im Bereich der Plattformökonomie – als Intermediär mittelfristig die Schnittstellen zwischen Datenerzeugern und Datennutzern zu besetzen und diese Dienstleistung zu monetarisieren. Mit Blick auf die gegenwärtige Kostenfreiheit sind entsprechende Angebote auch für Wissenschaftlerinnen und Wissenschaftler bis hin zu wissenschaftlichen Organisationen und Hochschulen attraktiv. Übersehen wird dabei häufig die in die Zukunft verschobene Gewinnerwartung über Gebührenmodelle oder vertragliche Verfügungs- und

Wertschöpfung
in der digitalen
Datenökonomie

¹⁶ Zu beobachten ist, dass aus dem generellen „Zur-Verfügung-Stellen“ ein gesplittertes Portfolio von Normen, wie Daten zu „teilen“ sind, entsteht.

¹⁷ Vgl. hierzu und zum Folgenden EFI (2016) – Jahresgutachten, B 3: Geschäftsmodelle der digitalen Wirtschaft, S. 60-79.

Verwertungsrechte. Nachdem eine kritische Masse von Nutzerinnen und Nutzern nicht mehr oder nur mit erheblichem Aufwand wechseln kann (Lock-in), ändern die Anbieter häufig die Gebührenmodelle. Ein prinzipiell möglicher Ausstieg aus dem Angebot kann dann mit hohen Wechselaufwendungen und/oder mit Transaktionskosten verbunden sein – vom einfachen Gewöhnungseffekt bis hin zu den materiellen Kosten einer Datenmigration.

Vermarktung von Daten über Forscherverhalten

Auch Informationen über das Verhalten von Nutzerinnen und Nutzern sind in diesem Zusammenhang kommerziell verwertbar. Vor diesem Hintergrund werden digitale, etwa auf Plattformen anfallende Daten bzw. bewusst nachverfolgte (getrackte) Datenspuren, die über Wissenschaftlerverhalten Auskunft geben können, zum neuen, zunehmend auch gesondert kommerziell relevanten Bestandteil von Geschäftsmodellen der internetbasierten Datenökonomie. So generieren z.B. einige für Nutzer kostenfreie Forschungsnetzwerke aus den Informationen, die Forschende dort einspeisen, um sich zu vernetzen, untereinander zu kommunizieren und Daten zu teilen, aggregierte Publikations- und Zitationsanalysen, die geldwert sind. Nutzerinformationen können hier vom Dienst als Innovationsquellen für Dritte genutzt werden. Das zugrundeliegende Geschäftsmodell wird auch als „**Erlebnisorientierte Crowdnutzung**“ beschrieben.¹⁸ Gleiches gilt für Servicepakete kommerzieller Verlage, die einerseits kostengünstige Tools für das Forschungsdaten- und Literaturmanagement bereitstellen, andererseits durch den Einsatz von Nachverfolgungsinstrumenten (Trackern) kommerziell verwertbare Informationen über den gesamten Forschungsprozess sammeln können. Dass auch Staaten Interesse daran haben könnten, Daten zum Wissenschaftlerverhalten auszuspähen, ist keine weit hergeholte Vermutung.

Geschäftsmodelle

Weitere Geschäftsmodelle, die sich auch in den von Wissenschaftlern genutzten Datendiensten wiederfinden, sind unter anderem:

- „**Kostenfreie Plattformen**“, die Dienste zahlungsfrei anbieten (Freemium), auf denen sie ggf. kostenpflichtige Zusatzangebote aufsetzen (Premium). Die Plattform nimmt hier häufig eine Koordinatorenrolle ein, in der sie eingespeiste Daten aus diversen Quellen in ein aggregiertes Produkt transformiert und anbietet.
- Durch „**zweiseitige Märkte**“ und die Berücksichtigung „differenzierter Nachfrage“ können unterschiedliche Nutzergruppen mit jeweils auf sie zugeschnittenen Produkten und differenzierten Gebührenmodellen bzw. Abonnementsystemen bedient werden.
- „**Partnerschaftsplattformen**“ lassen sich die Koordinatorenrolle vergüten, wenn beispielsweise „Kunden“ an Datenerzeuger vermittelt werden.

¹⁸ Ebd.

Diese und weitere Geschäftsmodelle treten in der internetbasierten Digitalwirtschaft häufig auch in Mischformen auf. Durch preisgünstige Angebote sind sie für die öffentlich finanzierte Wissenschaft (und die Ausschreibungslogik öffentlicher Beschaffung generell) attraktiv. Es mehren sich allerdings die Stimmen, die vor den Folgen eines flächendeckenden Lock-ins und des Einsatzes von Tracking-Methoden warnen, da dergleichen bereits mittelfristig die Souveränität des Wissenschaftssystems unterlaufen kann.¹⁹

Auch Wissenschaft kann zum Zweck der Verwertung am Ende eines Forschungsprozesses auf Grundlage von Daten und der mit den Daten erzielten Ergebnisse „Produkte“ erzeugen. Auch wenn diese dann faktisch monetarisiert werden, findet in der Regel auf Seiten der Wissenschaft keine „Vermarktung“ statt. Vielmehr kooperiert Wissenschaft hierzu mit unternehmerisch tätigen Intermediären oder Partnern. Diese können sich durchaus unterschiedlicher Geschäftsmodelle bedienen, die sich bei einer heuristischen Betrachtung zwischen den Polen eines Supply-Side-Modells und eines Demand-Side-Modells bewegen.

Datenprodukte und Intermediäre in der Wissenschaft

Beim Supply-Side-Modell vergüten die Datenerzeuger – also einzelne Wissenschaftler oder (i.d.R.) wissenschaftliche Einrichtungen – den Intermediär dafür, dass er ihre Daten archiviert und ggf. sogar qualitätsgesichert (kuratiert) oder mit Analyse-Services angereichert potentiellen Nutzern kostenfrei zur Verfügung stellt. Beim Demand-Side-Modell entstehen der „einliefernden“ Wissenschaft dagegen keine Kosten, diese erhebt der Intermediär bei den Nutzern.

Demand- und Supply-Modelle

In Deutschland sind auch „Ausgründungen“ aus der Wissenschaft ein typischer Weg, der von der Verwertung in die Vermarktung führt – wobei dann neue Unternehmen entstehen. Auch im Falle einer Verwertung von Datenprodukten verhält Wissenschaft sich bislang somit nicht „wirtschaftlich“. Vielmehr sind die Welten der (freien) wissenschaftsinternen Nutzung und diejenige der (auf Gewinnchancen ausgerichteten) Vermarktung durch geordnete Routinen sowie differente, interessengerechte Rollen getrennt.

Ausgründungen als Weg in die Vermarktung

¹⁹ Grundsätzlich dazu: WR (2021) – Impulse aus der COVID-19-Krise. Zum Tracking von Wissenschaft veröffentlichte der Ausschuss für Wissenschaftliche Bibliotheken und Informationssysteme der DFG ein Informationspapier, in dem er auf die Verschiebung des Geschäftsmodells von Verlagen hin zur Datenanalytik hinweist und für klare rechtliche Regulierung, hohe Transparenz und Mitbestimmung der Wissenschaft plädiert. Das Papier unterscheidet drei Typen der Verfahren zur Datengewinnung: Third Party Data durch Microtargeting, Bidstream Data sowie Verlagstrojaner. AWBI (2021) – Datentracking in der Wissenschaft, S. 5 ff., 9. Vgl. auch: Brembs/Foerstner (2020) – Überwachte Wissenschaftler; sowie die Initiative „Stop Tracking Science“ (stoptrackingscience.eu), zuletzt abgerufen am 2.06.2021. Der RfII behält sich vor, zu Einsatz und Folgen entsprechender Technologien im Forschungsprozess zu einem späteren Zeitpunkt Stellung zu beziehen.

Common Sense
als Grundlage
wissenschaftlichen
Datenteilens

Wissenschaftsinterne Nutzung im Sinne des freien Datenteilens unter Forscherinnen und Forschern war in diesem Zusammenhang bislang auch kaum durch rechtliche Regulierung als vielmehr durch einen breit akzeptierten Common Sense getragen, der wesentlicher Bestandteil einer „guten wissenschaftlichen Praxis“ ist. Aufgedeckte Verstöße gegen diesen Common Sense werden bislang in der Regel unterhalb der Schwelle juristischer Sanktionierung mit Reputationsentzug „bestraft“. Eine stärkere gesetzgeberische Regulierung des digitalen Teilens von Daten, die Forschung einschließt und Zugriffsrechte kommerzieller Interessenten stärkt, könnte künftig auch zu einer dichteren rechtlichen Durchdringung der wissenschaftlichen Gepflogenheiten führen.

1.3 ZU ZIELSTELLUNG UND STRUKTUR DES POSITIONSPAPIERS

Besseres Verständnis
für die praktizierten
Geschäftsmodelle
durch Fallbeispiel-
analyse

Mit einer Fallbeispielanalyse von existierenden Datendiensten hat der RfII zunächst den Wunsch verbunden, ein besseres Verständnis für die bereits praktizierte Nutzung und Verwertung von Forschungsdaten zu gewinnen. Von Interesse waren die *Zugangsregimes* der jeweiligen Dienste und ggf. auch ihre *Geschäftsmodelle* sowie deren Kompatibilität mit der EU-seitig geforderten „Politik des offenen Zugangs“. Auch sollten aus der Perspektive des Wissenschaftssystems heraus *Anforderungen* an die Gestaltung von Diensten und Zugangsregimes erarbeitet werden. Dies zu verstehen kann zu einer wissenschaftsadäquaten Ausgestaltung von Regelwerken („Policies“) und übergreifenden Strategien zur Datenpolitik in Deutschland und Europa ebenso beitragen, wie zur Weiterentwicklung von operativen Forschungsdatenstrategien auf der Ebene der Wissenschaftsorganisationen und Hochschulen. Dies ist vor allem dort bedeutsam, wo

- Dienste nicht in der Wissenschaft entwickelt wurden, sondern Wissenschaft sich externer, kommerziell ausgerichteter Dienste bedient;
- in der Wissenschaft entwickelte Dienste keine öffentliche Förderung (mehr) finden und z.B. durch Gebührenerhebung beim Zugang oder andere Formen monetärer Verwertung von Inhalten (Nutzung) oder der Beobachtung des Nutzerverhaltens (Tracking) ihren Erhalt sichern müssen;
- in der Wissenschaft entwickelte Dienste von kommerziellen Anbietern übernommen werden.

Limitierungen der
Analyse

Das Positionspapier fasst die Befunde aus der Sichtung von gut vierzig Datendiensten zusammen, die über verschiedene Disziplinen, Forschungsfelder und Domänen verteilt sind. Die Auswahl erhebt keinerlei Anspruch auf Repräsentativität oder Vollständigkeit der zu einzelnen Diensten erhobenen Informationen. Letztere wurden erhoben, soweit sie öffentlich im Rahmen einer internetbasierten Recherche zugänglich waren. Der Großteil der den Diensten zugrundeliegenden Geschäftsmodelle lässt sich auf dieser Basis nur pauschal beurteilen, da zu viele Informationen unbekannt bleiben oder über die einzelnen

Dienste hinweg nur sehr ungleichgewichtig erhoben werden konnten. Dies erschwert einen Vergleich, der den Standards einer robusten wissenschaftlichen Komparatistik genügen würde, lässt aber dennoch auf einer stark abstrahierenden Ebene Einschätzungen und Rückschlüsse zu, die aus Sicht des RfII für Wissenschaft und Wissenschaftspolitik einen Orientierungsnutzen beanspruchen können. Tabelle 1 und der Anhang geben einen Überblick über die insgesamt betrachteten Dienste. Dort, wo belastbare Informationen vorliegen, verweist der Text auf gekennzeichnete Praxisbeispiele, die der besseren Anschaulichkeit dienen. Bewertungen werden aufgrund der nicht repräsentativen Auswahl und der nicht vollständigen Informationslage zu den Diensten auf einem verallgemeinernden Niveau vorgenommen, nicht mit Bezug auf einzelne Beispiele.

Die folgenden Kapitel legen ihren Fokus auf die Frage, welche Dienste bereits im Bereich der Forschungsdaten und für die Forschung relevanter Daten entstanden sind (auch hier ohne Anspruch auf Vollständigkeit), in welchen Konstellationen sich diese Dienste entwickelt haben und welche Schlüsse sich aus der Verfasstheit der aktuellen Landschaft der Datendienste für die Nutzung und Verwertung der Daten ziehen lassen. Fragen der Vermarktung spielen dabei insofern eine Rolle, als einige dieser Dienste privatwirtschaftlich organisiert sind und mit ihren Geschäftsmodellen zumindest mittelfristig Erträge erwirtschaften müssen. Eine gewerbliche Verwendung der Daten wird eher am Rande behandelt und ist nicht zentraler Gegenstand der abschließenden Schlussfolgerungen.

Fokus auf
wissenschaftlich
relevante
Daten(dienste)

Kapitel 2 gibt zunächst einen Überblick, wie und aus welchen Gründen sich Datendienste in verschiedenen wissenschaftlichen und wissenschaftsnahen Domänen entwickelt haben. Der Ursprung der Entwicklung von Diensten und ihrer Zwecke aus den Paradigmen und arbeitsteiligen Forschungsprozessen verschiedener Domänen heraus kann allerdings die Vielgestaltigkeit von Betriebs- und Geschäftsmodellen zurzeit existierender Dienste nicht vollständig erklären. Dies zeigt sich insbesondere auch im Bereich von Datenbanken bzw. Repositorien, in denen wissenschaftlich erzeugte Daten unabhängig von deren fachlicher Herkunft und mit multiplen, aber wenig spezifischen (Nach-)Nutzungsmöglichkeiten archiviert werden: den so genannten Long-Tail-Daten.²⁰ Es bedarf demnach weiterer Analyse Dimensionen und Merkmalsausprägungen um die Funktionalität von Datendiensten für die Wissenschaft ebenso wie die wissenschaftliche Angemessenheit von deren Betriebs- und Geschäftsmodellen beurteilen zu können.

²⁰ Der Begriff leitet sich aus der Darstellung von Verteilungskurven für Forschungsdaten ab: Je nach betrachtetem Aspekt bzw. Variable auf der y-Achse weist ein Großteil der Datensätze nur geringe Umfänge, Homogenität, Nutzungszahlen, Einnahmen o.Ä. auf und bildet damit den „langen Schwanz“ (long tail) der Verteilungskurve. Zur Begriffsbestimmung vgl. genauer RfII (2016) – Leistung aus Vielfalt, Anhang A, Nr. 2.20.

Zu diesem Zweck gibt *Kapitel 3* Begriffe und Konzepte an die Hand, die einen domänenübergreifenden Vergleich zu Funktionen und Folgen von Diensten und ihren Geschäftsmodellen leisten können. In diesem Zusammenhang sind Merkmale und deren empirische Ausprägungen wie „Akteursrollen“ (z.B. Datenerzeuger oder Datennachfrager), im Rahmen der Dienste eingesetzte Technologien, Governance-Merkmale, Zugangsregimes und Fragen der (Re-) Finanzierung von Bedeutung.

Anschließend werden in *Kapitel 4* die verallgemeinerbaren Befunde der Fallbeispielanalyse sowohl unter Bezug auf die identifizierten Domänenspezifika, aber mehr noch mit Blick auf die übergreifenden Merkmale und deren Ausprägungen verglichen und diskutiert.

Aus dieser Diskussion leitet der RfII in einem letzten Schritt die in *Kapitel 5* formulierten wissenschaftspolitischen Empfehlungen an verschiedene Adressatenkreise ab.

Welche Dienste unter den oben dargestellten analytischen Einschränkungen im Einzelnen betrachtet wurden und welchen wissenschaftlichen Feldern bzw. Domänen sich diese in etwa zuordnen lassen, lässt sich der folgenden Tabelle 1 entnehmen. Kurzdarstellungen der betrachteten Dienste finden sich im Anhang B.

Tabelle 1: Betrachtete Dienste der Fallbeispielanalyse, geordnet nach Domänen

Domäne	Fallbeispiele
Erd- und Umweltwissenschaften	CERA – Langzeitarchiv des Deutschen Klimarechenzentrums COPERNICUS – Europäische Sentinel-Daten Google Earth Engine OpenStreetMap (OSM) PANGAEA – Data Publisher for Earth & Environmental Science Socioeconomic Data and Applications Center (SEDAC) World Stress Map (WSM)
Naturwissenschaften	Chemical Abstracts Service (CAS) ChemSpider DETERM Inorganic Crystal Structure Database (ICSD) Novel Materials Discovery Repository and Archive (NOMAD)
Lebenswissenschaften	Biobanken Datenintegrationszentren der Medizininformatik-Initiative Ontosight The Arabidopsis Information Resource (TAIR) TriNetX Worldwide Protein Data Bank
Sozial-, Verhaltens- und Wirtschaftswissenschaften	Forschungsdatenzentrum der Bundesagentur für Arbeit (FDZ BA im IAB) Forschungsdatenzentrum des Leibniz-Instituts für Bildungsverläufe (FDZ LifBI) Forschungsdatenzentrum der Statistischen Ämter des Bundes und der Länder (FDZ Bund, FDZ Länder) Forschungsdatenzentrum des Sozio-oekonomischen Panels (FDZ SOEP) PsychArchives – Disciplinary Repository for Psychological Science SowiDataNet datorium Statista
Geistes- und Kulturwissenschaften	Bavarikon – Internetportal Kultur und Wissensschätze Bayerns Beck-online Bildarchiv Foto Marburg Deutsche Digitale Bibliothek (DDB) Deutsche Fotothek Deutsches Referenzkorpus (DeReKo) Deutsches Textarchiv (DTA) Digitales Wörterbuch der Deutschen Sprache (DWDS) iDAI.objects arachne Juris.de – Das Rechtsportal openJur Prometheus – Das verteilte digitale Bildarchiv für Forschung und Lehre
Heterogene, sog. Long-Tail-Daten	Dryad Digital Repository EUDAT Collaborative Data Infrastructure (EUDAT CDI) Figshare Mendeley Data RADAR Research Data Repository Zenodo

2 ENTWICKLUNG IN UNTERSCHIEDLICHEN DOMÄNEN UND IM BEREICH GENERISCHER DATENDIENSTE

In diesem Kapitel wird ein Korpus von rund 40 Fallbeispielen zunächst wissenschaftlichen bzw. wissenschaftsnahen Feldern zugeordnet (siehe Tabelle 1). Da eine Untergliederung in wissenschaftliche Disziplinen wegen der zahlreichen Überschneidungen nicht trennscharf erfolgen kann, wählt die Darstellung das etwas gröbere Raster der fachlich-thematischen „Domänen“. Ziel der Analyse ist es, Anforderungen an die Gestaltung von Diensten und Zugangsregimes aus der Perspektive der Wissenschaft heraus deutlich zu machen und diese ggf. an Erwartungen aus Politik und Gesellschaft zu spiegeln. Ein erster Zugang über fachliche Domänen bietet sich an, weil dadurch am ehesten verständlich wird, welche wissenschaftlichen Bedürfnisse nach Datendiensten und nach spezifischen Ausprägungen von Betriebs- und Geschäftsmodellen die heutigen Problemstellungen im Bereich der Nutzung und Verwertung begünstigt haben.²¹

Datendienste zwischen „Normalwissenschaft“ und multiparadigmatischen Zugängen

Es zeigt sich, dass gerade in solchen wissenschaftlichen Feldern ein fest etabliertes Segment von langlebigen und mit Reputation aufgeladenen Diensten entstanden ist, in denen erstens „Normalwissenschaft“ im Sinne eines weitgehend monoparadigmatischen, also methodisch recht einheitlichen „Puzzle Solving“ betrieben wird, und in denen sich zweitens hierfür früh ein hoher Bedarf an internationaler Kooperation und Arbeitsteilung ergeben hat. Gleichwohl haben viele wissenschaftlich anerkannte Dienste auch in diesem Bereich keine verstetigte institutionelle Absicherung. Dort, wo Schnittstellen mit kommerziellen Verwertungsinteressen vorgezeichnet sind, finden sich in diesen Feldern zudem – auch im Rahmen wissenschaftlicher Fachgemeinschaften – Dienste, die sowohl innerwissenschaftliche Nutzung als auch außerwissenschaftliche Verwendung durch ausdifferenzierte Geschäftsmodelle miteinander kompatibel zu machen versuchen. Auch zur Erschließung von alltagsweltlichen²² Daten für Forschungszwecke haben mehrere Domänen Formate entwickelt.

Schwieriger scheint sich ein nachhaltiger und qualitätsgesicherter Aufbau von Diensten dort zu gestalten, wo sich wissenschaftliche Domänen durch multi-

²¹ Der Wissenschaftsrat weist darauf hin, dass der Entwicklungsstand von Standards und Strukturen in jeweiligen Fachgemeinschaften sich unterschiedlich schnell, je nach Akzeptanz und Bedarf, verändert: Astronomie, Hochenergiephysik und Lebenswissenschaften gehören zu den Vorreitern, andere Disziplinen dagegen benötigen für ihre Erkenntnissuche und Ergebnisqualität weniger oder auch gar keine datenintensiven Methoden. WR (2020) – Zum Wandel in den Wissenschaften, S. 13. Das hat Auswirkungen auf die Ansprüche, die verschiedene Disziplinen, Forschungsfelder und Domänen an Datendienste stellen bzw. – vom Dienst her gedacht – welche Bedürfnisse bedient werden müssen. Dies berührt insbesondere Fragen der Qualitätssicherung von Daten und ihrer Zugänglichkeit (u.a. auch Fragen der Zumutbarkeit von Gebührenmodellen und Bezahlschranken).

²² Zum Beispiel aus öffentlich betriebenen Umweltmessenanlagen, der Krankenversorgung, dem öffentlichen Nahverkehr, den sozialen Netzwerken usw.

paradigmatische Zugänge auszeichnen, wo Forschung, durch unterschiedliche methodische Zugänge (Schulen) geprägt, die Methoden mischt oder wechselt, und wo weniger klare Zwecke für die Nutzung und Verwertung von Forschungsdaten gegeben sind. In solchen Feldern zeigt sich eine oft uneinheitliche Gemengelage aus „kleinen“ projektgeförderten Diensten mit teils geringerer Standardisierung und Nachhaltigkeit des Leistungsspektrums. Forschungsparadigmen und Methoden können dabei konkurrieren – und zwar in anerkannter Weise, da Wissen als perspektivisch und modellabhängig gilt. Prozesse der zunehmenden Kooperation und Bildung eigener koordinierender Akteure (wie z.B. des RatSWD in den Sozialwissenschaften) haben hier in Teilbereichen für Abhilfe gesorgt und Fragen der Nutzung und Verwertung teilweise verbindlich regeln können.

Die genannten Phänomene – etablierte (auch große) eigene, teil-kommerzielle oder kommerzielle sowie (oft kleine) heterogene, unverbundene wie auch koordiniert aufgebaute Datenbestände und Datendienste – finden sich letztlich in allen Domänen mehr oder weniger ausgeprägt wieder. Der große Bereich der teils fachlich gar nicht klar geprägten (einerseits sehr spezialisierten, andererseits potentiell übergreifend interessanten) Long-Tail-Daten, dessen unübersichtlicher Pool sich aus allen im Folgenden dargestellten Domänen speist, kommt hinzu. Er ist allein durch die Adresse, unter der er aufbewahrt wird (etwa auf großen, unspezifischen Plattformen oder Forschungsdatenrepositorien), definiert und aufgrund seiner Heterogenität auch für Fragen der Nutzung und Verwertung am schwierigsten zu greifen. Hier sind zahlreiche generische Publikations- und Archivierungsdienste etabliert, die sich in Art und Umfang am wissenschaftlichen Publikationssystem orientieren, speziell den Bibliothekskatalogen und den Fachzeitschriften. Long-Tail-Repositorien sind auf einige begrenzte, eher innerwissenschaftliche Nachnutzungsszenarien optimiert und folgen ansonsten generellen informationswissenschaftlichen Standards (etwa „DOI“). Neben einigen übergreifenden, auch kommerziellen Diensten, wird für Long-Tail-Daten auch eine unüberschaubare Zahl lokaler Repositorien an Universitäten und Forschungseinrichtungen betrieben. Entsprechend divers gestalten sich hier nicht nur Betriebs- und Geschäftsmodelle, sondern auch Ansprüche an Langzeitarchivierung und Qualitätssicherung.

Heterogenität der Nutzung und Verwertung bei generischen Publikations- und Archivierungsdiensten

2.1 ERD- UND UMWELTWISSENSCHAFTEN

Die Domäne der Erd- und Umweltwissenschaften ist durch ein hohes Maß an globaler Zusammenarbeit und durch eine lange Tradition der Sammlung und Bereitstellung von Daten gekennzeichnet, sowohl was die Forschung als auch was das Handeln von staatlichen Behörden betrifft. Für Standardisierungsinitiativen lassen sich zahlreiche Beispiele finden. Daten aus der Wissenschaft sind eng verbunden mit Geodaten, die von Behörden und Ämtern oder im Rahmen staatlicher Programme erzeugt werden (etwa große Satellitenmissionen). Da diese Daten

Standardisierungsinitiativen und transnationale Betreibernetzwerke über die Wissenschaft hinaus

weltweit aus allen Sektoren nachgefragt werden, sind grenzüberschreitende Zusammenarbeit sowie transnationale Betreibernetzwerke und Programme fest etabliert und teils hochformell durch Staatsverträge abgesichert.

Großvolumige dynamische Daten

In einigen Fallbeispielen kommen neuere Entwicklungen im Bereich der Plattformtechnologien zur Anwendung, und zwar speziell für die Verwertung und Vermarktung großvolumiger dynamischer Daten, die laufend aus Messgeräten verschiedener Art übermittelt werden – unter anderem aus Geräten der Landwirtschaft, Luftreinheitsüberwachung, Wetterbeobachtung etc. Paradigmatisch dafür wurden in der Fallbeispielanalyse Dienste für satellitengestützte Erdbeobachtungsdaten betrachtet. Die Erdbeobachtung insgesamt erzeugt weltweit mit die größten Datenmengen. Das stets wachsende Datenvolumen erfordert hohe Rechen- und Speicherkapazitäten sowie umfassende Transformationen, um Daten für Anwender nutzbar zu machen.

Umfangreiche offene Datenbestände und lebhafter Markt für Geoinformationsdienste

Da Daten oftmals offen und für gewerbliche und nichtgewerbliche Nutzung gleichermaßen zur Verfügung gestellt werden, können ebenso große kommerzielle Betreiber auf umfangreiche Datenbestände für eine Vermarktung zurückgreifen. In der Domäne hat sich zudem ein lebhafter Markt für sogenannte Geoinformationsdienste²³ und damit verbundene Dienstleistungen entwickelt. Hier zeigen sich positive Effekte von Open Data auf die Datenwirtschaft. Diese werden seitens der Politik als Vorbild für die Gestaltung von Informationsinfrastrukturen in der Wissenschaft (zum Beispiel der EOSC) herangezogen, von denen man sich eine ähnliche Wertschöpfung erhofft. Das Engagement staatlicher Akteure jenseits des Wissenschaftsressorts oder von Großunternehmen trägt allerdings auch schon bei der Entstehung der Dienste zur Schaffung der Verwertungsmöglichkeiten bei.

Praxisbeispiel 1

Im Fall des europäischen COPERNICUS Programms werden die Daten von der zuständigen Weltraumagentur ESA verwertet, die diese über einen „Open Access Hub“ offen und kostenfrei zur Verfügung stellt. Darüber hinaus fördert die EU die Vermarktung der Daten über technologisch anspruchsvolle Plattform-Dienste, deren Betreiber sich auf konkrete Verwendungen in verschiedenen Branchen und Sektoren spezialisieren und diese ggf. auch mit Daten anderer Anbieter kombinieren, zum Beispiel aus den Bereichen Verkehr, Klima oder Landwirtschaft.²⁴

²³ Informationssysteme zur Erfassung, Bearbeitung, Organisation, Analyse und Präsentation räumlicher Daten.

²⁴ Data and Information Access Services, DIAS (copernicus.eu/de/datenzugriff/dias), zuletzt abgerufen am 2.06.2021.

Die Plattformtechnologien (Infrastructure-as-a-Service, Software-as-a-Service) sollen eine Nutzung der enormen Datenmengen unabhängig von der eigenen IT-Ausstattung ermöglichen. Als Intermediäre agieren hier zum Teil die großen europäischen Telekommunikationsunternehmen, aber auch kleinere Unternehmen, die Applikationen für die Analyse und Visualisierung von Daten entwickeln und diese über die Plattformen vermarkten. Wissenschaft entwickelt selbst auch Applikationen, tritt aber primär als Datennachfrager auf. Mit der COPERNICUS Data and Exploitation Plattform (CODE-DE) am Deutschen Zentrum für Luft- und Raumfahrt (DLR) wird in Deutschland zum Beispiel ein spezifisch wissenschaftsorientierter Zugang zu den europäischen Satellitendaten geschaffen. Mit der technischen Umsetzung wurden in einem Ausschreibungsverfahren fachnahe Dienstleistungsunternehmen beauftragt.

Die wissenschaftliche Forschung greift unter anderem auf die oben beschriebenen staatlicherseits bereitgestellten Daten zurück. Es werden aber auch in großem Umfang eigene Forschungsdaten erhoben. Die Datenkultur in den Geowissenschaften kann in Teildisziplinen als gut entwickelt gelten. Einen Beitrag hierzu leistet wohl auch die fortgeschrittene Digitalisierung der Forschungsprozesse selbst.²⁵ Mit der Geoinformatik hat sich zudem eine Teildisziplin herausgebildet, die spezielle Anwendungen für die Domäne entwickelt. Einer Erhebung zufolge gibt es über 100 primär innerwissenschaftlich organisierte Datendienste, allerdings nicht alle nachhaltig finanziert. Gut etablierte Dienste zur Veröffentlichung und Archivierung von Daten aus der Projektforschung sind zum Beispiel der *PANGAEA Data Publisher* oder *CERA*, das Langzeitarchiv des Deutschen Klimarechenzentrums. Hinzu kommen transnationale Datensammlungen mit weltweit anerkannten Qualitätsmanagementkonzepten: Das *Socioeconomic Data and Applications Center (SEDAC)* der NASA trägt für seine kuratierte Datensammlung raumbezogene Daten aus der Forschung zu Mensch-Umwelt-Interaktionen zusammen, die entsprechend einer Data Acquisition Policy vor der Aufnahme begutachtet werden. Die *World Stress Map* bereitet Daten aus Kernbohrungen auf und verwendet ein „quality ranking scheme“ für jeden Stress-Indikator, um die Aussagekraft aller Datensätze vergleichbar zu halten. Diese qualitativ hochwertigen Daten werden auch von Unternehmen im Berg- und Hochbau nachgefragt.

Etablierte Dienste
zur Veröffentlichung
und Archivierung
von Daten

Eine Verwertung von Forschungsdaten durch ihre Veröffentlichung mit enger technischer Anbindung an das Publikationssystem (Verknüpfung von Datensätzen mit anderen digitalen Objekten wie wissenschaftlicher Literatur) ist relativ gängig.

²⁵ Bernard/Mäs (2018) – e-science in den Geo- und Umweltwissenschaften.

Zudem birgt die Domäne der Erd- und Umweltforschung zahlreiche Beispiele für ein Engagement der Zivilgesellschaft. Dies äußert sich etwa in Kooperationen zur Datenerzeugung mit der Wissenschaft – so in den letzten Jahren in den Citizen-Science-Projekten der Helmholtz-Gemeinschaft. Mit dem schon 2004 gegründeten Projekt *OpenStreetMap* ist zudem ein Datendienst etabliert, der (vergleichbar dem Wikipedia-Ansatz) kollektiv verantwortet wird – von der Datenerzeugung über die Kuratierung bis hin zur Finanzierung. Aktiv sind primär zivilgesellschaftliche Akteure, aber auch Behörden, individuelle Wissenschaftler und Wissenschaftlerinnen sowie wissenschaftliche Einrichtungen unterstützen das Projekt. Auch aufgrund dieses zivilgesellschaftlichen Engagements kann die Domäne Erd- und Umweltwissenschaften als beispielhaft für die Umsetzung des „Open Science“-Paradigmas gelten.

2.2 NATURWISSENSCHAFTEN

In den Naturwissenschaften ist die Frage nach einer außerwissenschaftlichen Verwertbarkeit und Vermarktbarkeit von Daten je nach Fach und Forschungsfeld differenziert zu betrachten. Daten in der Chemie und in den Materialwissenschaften haben beispielsweise einen hohen Marktwert. Fachliche Datenangebote richten sich deshalb erkennbar gleichermaßen an wissenschaftliche und an gewerbliche Nachfrager. Wissenschaftliche Akteure verwerten in dieser Domäne darüber hinaus die von ihnen gewonnenen Daten auch selbst kommerziell.

Chemische und materialwissenschaftliche Daten werden sowohl analog wie auch digital (Laborexperimente, Modellierung) erzeugt. Sie sind heterogen hinsichtlich des Inhalts, des Umfangs und der Datenformate, die nicht selten proprietär (durch von Unternehmen vertriebene Produkte) entstehen. Ihre (fehlende) Interoperabilität ist damit für die Verwertung in der Forschung ein häufiges Hindernis, die Domäne hat mit dem Phänomen eines „Blackboxing“ in Datenerzeugung oder -prozessierung zu kämpfen.²⁶ Als organisatorisches Rückgrat für die Informationsinfrastrukturen dienen Glossare, Terminologien und andere Standards, die bereits seit mehr als hundert Jahren gepflegt und angewendet werden.²⁷ Daten aus der Individualforschung werden üblicherweise als Supplementary Materials mit detaillierten, experimentellen Daten in Tabellen, Strukturformeln und Abbildungen von Spektren veröffentlicht, jedoch sind diese Inhalte oftmals weder direkt referenzierbar noch maschinenlesbar.

²⁶ Vgl. zum Problem „Blackboxing“ RfII (2019) – Herausforderung Datenqualität, S. 32, 49, 68, 87.

²⁷ Dabei spielt zum Beispiel die Fachgesellschaft IUPAC (International Union of Pure and Applied Chemistry) eine zentrale Rolle. Forschende verweisen in Fachpublikationen regelhaft auf die verwendeten IUPAC-Standards.

Typisch sind fernerhin aufwändige fachliche Hochdurchsatz-Analysen und der Abgleich von Daten (z.B. Strukturformeln). Deshalb finden sich einerseits redaktionell kuratierte, systematisch um Daten aus der aktuellen Forschung ergänzte Referenzkorpora, andererseits ist hier für die Forschung – und dies wird von der Fachgemeinschaft auch ausdrücklich gefordert – die Veröffentlichung von Negativbefunden (Null-Results) besonders wichtig, bleibt aber bislang ein Desiderat.²⁸

Negativbefunde sind werthaltige Daten

In der Fallbeispielanalyse wurden Datendienste zweier großer chemischer Fachgesellschaften betrachtet, die Tausende Mitglieder haben und an der Schnittstelle Wissenschaft/Wirtschaft agieren. Sie geben die Referenzkorpora für das Fach heraus, organisieren die wissenschaftliche Kuratierung der Daten, entwickeln und betreiben internationale Datendienste, die sie entweder selbst (gewinnbringend) vermarkten oder sich dafür kommerzieller oder wissenschaftsinterner Intermediäre als technische Dienstleister bzw. als Vertriebspartner bedienen. So hat die American Chemical Society (ACS) ihren seit 1907 (seit 1966 als eine digitale Datenbank) bestehenden *Chemical Abstracts Service (CAS)* um neue Inhalte und Werkzeuge bzw. Plattformtechnologien erweitert und Premium-Modelle entwickelt, die wissenschaftliche Expertise auch für Beratung bei unternehmerischen Entscheidungen vermarktet. Die Datenprodukte der ACS werden weltweit an Industriekunden vertrieben, für die Wissenschaft stehen preislich und im Umfang reduzierte Varianten zur Verfügung. *ChemSpider*, eine seitens der britischen Royal Society of Chemistry betriebene, umfangreiche Datenbank mit chemischen Strukturen und Texten, ist hingegen ein kostenfreier und kollaborativ kuratierter Dienst. Auch Werbeeinnahmen tragen hier zur Finanzierung bei.

Fachgesellschaften als zentrale Akteure

Darüber hinaus existieren zahlreiche Datendienste bzw. Datenbanken in Teildisziplinen wie Thermophysik, computergetriebene Materialwissenschaften oder Kristallografie. Das Engagement der Teildisziplinen für die Kuratierung der Daten zeigt sich unter anderem in der Entwicklung von Standards. Das von Cambridge Structural Database (CSD) des Cambridge Crystallographic Data Centre (CCDC) seit 1965 entwickelte Format (CIF) ist mittlerweile als Standardformat für Kristallstrukturdaten etabliert. Was die Anbieterseite betrifft, so finden sich sowohl Public-Private-Partnerships als auch Ausgründungen aus der Wissenschaft, etwa zur Verstetigung von teilweise langjährig drittmittelgestützten Aktivitäten. Teilaufgaben wie Datenerzeugung und Kuratierung verblieben auch in diesen Konstrukten zum Teil in der Verantwortung der Wissenschaft. Als typisch für die Entstehungsgeschichte vieler Dienste lassen sich Wechsel

Datenkuratierung und Standards in Teildisziplinen

²⁸ Vgl. Pappenberger (2016) – bwFDM-Communities. – Negativbefunde könnten in der Domäne Chemie und Materialwissenschaften im Übrigen auch vermarktet werden: Sie sind für Nachfrager aus der forschenden Wirtschaft von hoher Bedeutung, da sie die Effizienz der eigenen Forschung steigern.

der Betreiberkonstellationen bzw. der Governance und des Geschäftsmodells feststellen. Bei kommerzialisierten Diensten werden der Wissenschaft in der Regel Sonderkonditionen eingeräumt. Der Zugang kann dabei allerdings auf Grundfunktionen beschränkt sein. Neben traditionellen Datenbanken findet sich auch der Einsatz moderner Plattformtechnologien.

Praxisbeispiel 2

Das *Novel Materials Discovery Repository and Archive (NOMAD)* ist ein Beispiel für einen Dienst, der eine Teildisziplin im gesamten Forschungsprozess unterstützen soll. In der letzten Dekade von Forschungseinrichtungen aus EU-Fördermitteln unter Beteiligung der Industrie aufgebaut, erweitert um Software-Werkzeuge (Artificial Intelligence Toolkit) und um eine Knowledge-Bases (Materials Encyclopedia), steht der Datendienst den Datenerzeugern und Nachfragern aus den computergestützten Materialwissenschaften kostenfrei zur Verfügung, harrt aber noch der Verstetigung. Aktueller Träger ist FAIR-DI e.V. mit Sitz in Berlin.

Anders als in den oben genannten Beispielen stellt sich die Lage in der Astronomie und Astrophysik dar, in denen es seit jeher um die sprichwörtlichen „astronomischen“ Datenmengen geht. Bereits in den 1960er Jahren haben sich Astronomen international vernetzt und in den 1970er Jahren begonnen, international einheitliche Datenformate für astronomische Bilder, Spektren und Zeitreihen zu definieren. Um das Jahr 2000 bildete sich bottom-up aus dem Feld heraus die *International Virtual Observatory Alliance (IVOA)*, die versucht, alle astronomischen Daten und Datensätze weltweit jeder Person zugänglich zu machen. Die Vision ist „the universe at your fingertips“. Ziel ist es dabei, weltweit gemeinsame Datenformate und -protokolle zu definieren, um interoperable Datendienste in der Astronomie und die Erstellung von Werkzeugen zu ihrer Nutzung zu schaffen. „Interoperabel“ heißt in diesem Zusammenhang, dass dank international vereinbarter bzw. derzeit entwickelter Standards die verschiedenen Komponenten einer Dateninfrastruktur in der Astronomie so zusammenpassen, dass sie je nach Aufgabenstellung mit minimalem Aufwand zu leistungsfähigen Wissenschafts-Workflows kombiniert werden können. Die Nutzer haben so einen flexiblen Baukasten aus Diensten, Instrumenten, Client-Programmen, Bibliotheken und eigenen Dateien zur Verfügung, in dem alle Komponenten aufeinander abgestimmt sind. Die IVOA versucht, die internationalen Aktivitäten auf diesem Gebiet zu koordinieren und einen Großteil der in der Astronomie verwendeten Interoperabilitätsstandards zu entwickeln. Die deutsche IVOA-Partnerorganisation ist das *German Astrophysical Virtual Observatory (GAVO)*.

Hier seien nur die zwei populärsten und weltweit von allen in der Astronomie und Astrophysik tätigen Forscherinnen und Forschern genutzten Daten- und Literaturdienste genannt:

Praxisbeispiele 3

Das *SAO/NASA Astrophysical Data System (ADS)* ist beim Harvard-Smithsonian Center for Astrophysics im Cambridge/Mass angesiedelt und wird größtenteils von der NASA finanziert. Ursprünglich als Publikationsdatenbank konzipiert, enthält das ADS inzwischen quasi die komplette astrophysikalische Literatur (über 15 Millionen Einträge) und bietet ganz verschiedenartige weitere Analysemöglichkeiten, inklusive Verlinkung zu vielen Originaldatenkatalogen und -sätzen.

Das *Centre de Données astronomiques de Strasbourg (CDS)* ermöglicht es mit den drei Komponenten *Simbad*, *VizieR* und *Aladin*, interoperabel mit fast allen größeren (und vielen kleinen) Datensätzen und Katalogen der Astronomie zu arbeiten. Dabei bietet *Simbad* Informationen über astronomische Objekte, die in wissenschaftlichen Artikeln veröffentlicht wurden, *VizieR* ist quasi eine Bibliothek astronomischer Kataloge und Tabellen, und *Aladin* ist ein Tool zum interaktiven Visualisieren von astronomischen Daten, Katalogen und Durchmusterungen.

Sowohl das ADS wie auch das CDS sind weltweit kostenlos nutzbar und werden von professionellen Forscherinnen und Forschern ebenso wie von Amateur-Astronomen und „Citizen Scientists“ sehr intensiv genutzt.

Astronomische Daten haben gegenüber wissenschaftlichen Daten aus anderen naturwissenschaftlichen Feldern, etwa der oben angeführten Chemie oder den Materialwissenschaften, keinen direkten wirtschaftlichen Nutzen. Die Dateninfrastrukturen in diesem Bereich werden innerwissenschaftlich aufgebaut und getragen.

2.3 LEBENSWISSENSCHAFTEN

In der Domäne der Lebens- (oder Bio-)wissenschaften sind die Datennutzung wie auch die Forschung selbst von einem sehr hohen Grad internationaler Vernetzung geprägt. Ähnlich wie in den Erd- und Umweltwissenschaften finden sich zahlreiche Standardisierungsinitiativen und Dienste, sowohl für digitale als auch für analoge Objekte.²⁹ Mit der Bio- bzw. Medizininformatik haben sich Teildisziplinen herausgebildet, die spezialisierte Anwendungen für die Praxis entwickeln. Die Datenkultur erscheint somit in verschiedenen Teilbereichen (Biomedizin, Molekularbiologie, Biodiversität usw.) recht fortgeschritten. Marktwert bzw. intensiven Wettbewerb um neue Erkenntnisse stimulieren Industriekooperationen, limitieren dadurch aber auch teilweise die uneingeschränkte Bereitschaft zum Datenteilen. In begrenztem Maße gilt dies auch für die Forschung an Nutzpflanzen und in der Veterinärmedizin.

Standardisierungsinitiativen und Dienste für digitale und analoge Objekte

²⁹ Die Registry of Research Data Repositories verzeichnet knapp über 2.000 als „disziplinär“ ausgewiesene Repositorien (re3data.org, zuletzt abgerufen am 2.06.2021).

Eine Trennlinie bei den Diensten verläuft oftmals zwischen humanmedizinischer Forschung und Daten aus der Forschung an anderen Organismen (Tiere, Pflanzen, Mikroorganismen).

**Mitwirken von
„Datensubjekten“
in der Humanmedizin**

In der Humanmedizin ist das Mitwirken der sog. „Datensubjekte“, also der Patientinnen und Patienten, unerlässlich (vgl. 3.2.1). Daten entstehen in der klinischen Versorgung (auch als „Routinedaten“ bezeichnet), bei der Nutzung intelligenter „personal medical devices“ wie Blutzuckermessgeräte, sowie in der klinischen Forschung (Studiendaten von Probandinnen und Probanden). Gemeinsam sind ihnen die Personenbeziehbarkeit und eine hohe Schutzstufe, weshalb ähnlich wie in Teilen der Sozialwissenschaften besondere Anforderungen bei Archivierung und Zugang bestehen. Datenerzeuger und -nachfrager kommen aus Einrichtungen der Gesundheitsversorgung, forschenden Unternehmen, Universitäten, außeruniversitären Forschungseinrichtungen (inkl. Ressortforschung) und Behörden. Die Daten sind zudem von hohem kommerziellem Interesse. Zur Erschließung der (analogen und digitalen) Daten aus der Humanmedizin gibt es dementsprechend ausgesprochen vielfältige Initiativen.

Praxisbeispiel 4

Die seit 2016 laufende Medizininformatik-Initiative des BMBF zielt darauf ab, Routinedaten der über 30 deutschen Universitätsklinika strukturiert zu erschließen. Dazu werden an den Kliniken „Datenintegrationszentren“ gegründet, die in den angeschlossenen Einrichtungen Daten aus der Krankenversorgung nach einem gemeinsam vereinbarten Schema („Kerndatensatz“) zugänglich machen. Der Nutzung der Daten für Forschungszwecke müssen Patienten ausdrücklich zustimmen. Ein Zugriff wird auf Antrag gewährt. Die Bereitstellung und Kuratierung von Daten ist Teil des institutionellen Mandats.

In der humanmedizinischen Grundlagenforschung sind die Deutschen Zentren für Gesundheitsforschung nach einer Evaluation des Wissenschaftsrats 2017 aufgefordert, die Forschungsdaten des Netzwerks für eine Nachnutzung besser zu erschließen.³⁰ Die Erschließung von Patientendaten für Behandler ist Ziel der „elektronischen Patientenakte“, die unter Federführung des Bundesgesundheitsministeriums zum 1. Januar 2021 eingeführt wurde. Für die Zusammenführung von Daten zu Vorerkrankungen und Medikationen von Patientinnen und Patienten wird eine eigene Infrastruktur – die Telematikinfrastruktur Gematik – aufgesetzt, an der alle maßgeblichen Akteure des Gesundheitssystems beteiligt sind. Zudem wird ein komplexer rechtlicher Regelungsrahmen aus mehreren ineinandergreifenden Gesetzen geschaffen. Als weitere Initiative ist das seit 2016 im Aufbau befindliche Netzwerk der Biodatenbanken zu nennen.

³⁰ WR (2017) – Weiterentwicklung der Deutschen Zentren der Gesundheitsforschung, S. 63.

Ein Plattformdienst zur Aggregation klinischer Studiendaten ist *TriNetX* (USA). *TriNetX* tritt als Netzwerk klinischer Studienzentren auf, das Kontakte zu forschenden Pharmaunternehmen herstellt und es technisch ermöglicht, Daten auf einer gemeinsamen Plattform bedarfsbezogen zu teilen (Datenmarktplatz, clubartige Nutzergemeinschaft). Daten werden im großen Stil weltweit akquiriert („Big Data“).

In der Molekularbiologie wurden in den letzten 15 Jahren zentrale Infrastrukturen über internationale Abkommen und Vereinbarungen zwischen Regierungen oder zwischen Förderern abgesichert. Zu nennen ist z.B. das transnationale, im Wesentlichen europäische Netzwerk ELIXIR, das zentrale Datendienste in den Lebenswissenschaften pflegt.³¹ Internationale Datenarchive für experimentelle Daten sind gut etabliert. Seit den 1990er Jahren wurden weltweit sog. „Deposit Databases“ und „Archival Data Repositories“ für bestimmte Typen von Daten, wie Nukleotid- oder Proteinsequenzen aufgebaut. Auch „hybride“ Repositorien, also eine Kombination von physischen Objekten und ergänzenden Daten (Sequenzen, Protokolle usw.) sind zu finden.³² In einigen experimentellen Teilbereichen wie der Genomforschung ist die Abgabe von Daten an eines der internationalen Datenarchive professionelle Pflicht. Als Dienste-Betreiber wirkt hauptsächlich die öffentliche Hand, nicht selten in Form von internationalen Konsortien, wie im Fall der *Worldwide Protein Data Bank*.

Internationale
Abkommen und
Vereinbarungen
zwischen Regierungen
und zwischen
Förderern

Das Netzwerk ELIXIR bemüht sich auch um die Definition und den Erhalt eines grundständigen Service-Portfolios für die Lebenswissenschaften und hat dazu Studien und Evaluationen durchgeführt. Bei den „Core Data Resources“ werden neben den oben dargestellten Datenarchiven auch sogenannte „Knowledge Databases“ unterschieden, d.h. besonders kuratierte Datenbestände für einen spezialisierten Nutzerkreis. Ein Beispiel hierfür ist *The Arabidopsis Information Resource (TAIR)*. Der Dienst aggregiert Informationen zu einem häufig verwendeten pflanzlichen Modellorganismus für einen Anwenderkreis aus Wissenschaft und Industrie. Dazu werden teils automatisiert neue Sequenzinformationen aus der wissenschaftlichen Literatur und den molekularbiologischen Forschungsdaten erhoben und durch ein Kuratoren-Team ausgewertet. Solche „Model Organism Databases“ bilden einen eigenen Datentypus, dessen Optimierung sich ein eigener Verband verschrieben hat, die „Alliance of Genome Resources“.

Kuratierung als
Aufgabe von
Netzwerken

³¹ ELIXIR findet sich als europäische Forschungsinfrastruktur seit 2006 auf der ESFRI Roadmap; es wird finanziert aus Zuwendungen der Mitglieder, hinzu kommen flankierende Projekte und Spenden. Einige wenige ELIXIR-Nationalknoten kooperieren mit der Wirtschaft. Die Nutzung der bereitgestellten Inhalte ist kostenfrei. Ob weitere Leistungen (v.a. zu begrenzten Ressourcen wie HPC, Cloud oder Beratung) für die Nutzerinnen und Nutzer kostenpflichtig sind, variiert je nach nationalen Zuwendungen und Bestimmungen der beteiligten Mitgliedstaaten.

³² Zum Beispiel die Deutsche Sammlung von Mikroorganismen und Zellkulturen (eine Leibniz-Einrichtung) oder Addgene, eine Non-Profit-Organisation, die qualitätsgeprüfte DNA-Materialien vertreibt. Sie waren allerdings nicht Gegenstand der Analyse.

Das Unternehmen *Ontosight* nutzt im großen Stil Daten aus öffentlich finanzierten Infrastrukturen wie Patent- und Sequenzdatenbanken, und kombiniert sie mit Informationen aus öffentlich zugänglichen Literaturdatenbanken. Mithilfe von KI-Anwendungen können auf Kundenanfrage umfangreiche Recherchen auf dem Datenkorpus durchgeführt werden. Der Service ist auf die Bedarfe eines vorwiegend industriellen Nutzerkreises abgestimmt, für die wissenschaftliche Nutzung werden Sonderkonditionen gewährt.

Solch eine gewerbliche Nutzung ist international kein Einzelfall. So beschreibt das Netzwerk ELIXIR eine „Welle“ von Gründungen kleiner und mittlerer Unternehmen, die auf Basis offener Daten aus den lebenswissenschaftlichen Datenarchiven arbeiten.³³

Vermarktung
bzw. gewerbliche
Verwendung
offener Daten

Die Geschäftsfeldentwicklung erfolgt im Bereich der gewerblichen Nutzung von Biodaten inzwischen nicht mehr nur durch große Biochemie-Konzerne, sondern auch durch KMU. Sie erscheint paradigmatisch für die Erwartungen, die seitens der Innovationspolitik an offene Daten gestellt werden. Die ELIXIR-Studie weist aber darauf hin, dass eine solche Vermarktung bzw. gewerbliche Verwendung offener Daten nur bei freiem Zugang möglich sei und einen hohen Anspruch an die Qualität der Daten stellt, die gut kuratiert und nachhaltig verwaltet sein müssen.

2.4 SOZIAL-, VERHALTENS- UND WIRTSCHAFTSWISSENSCHAFTEN

Personenbezogene
Daten und
Längsschnittstudien

Datendienste in den Sozial-, Verhaltens- und Wirtschaftswissenschaften haben besondere Herausforderungen zu meistern: die Daten sind in besonderer Weise schutzwürdig, sollen aber zugleich für Längsschnittstudien interoperabel³⁴ und langzeitarchiviert bereitgestellt werden. Die personenbeziehbaren Daten in diesem Bereich werden von Forschungseinrichtungen durch Umfragen und Studien sowie von Ämtern und Behörden zum Beispiel über statistische Erhebungen erzeugt. Wegen des schutzwürdigen Charakters ist ein kollegiales Teilen solcher Daten in der und für die Wissenschaft naturgemäß nicht ohne Einschränkungen möglich, Zugangskonzepte sind daher notwendig. Nicht zuletzt ist Datenerzeugung mit Belastungen für Datensubjekte verbunden, was

³³ ELIXIR (2017) – Public Data Resources, S. 5. Eine gewerbliche Nutzung kann dann wie folgt aussehen: Anpassung (und Beratung): offene Daten werden den Zwecken des Kunden entsprechend angereichert („customising“); Aggregation (und Beratung): Offene Daten werden mit geschlossenen Daten des Diensteanbieters und ggf. mit den Daten des Kunden integriert („aggregating“); Ermöglichung: offene Daten und offene Software werden durchsucht und nach individuellen Bedarfen des Kunden zusammengestellt („enabling“).

³⁴ In den Sozialwissenschaften ist die Nutzung proprietärer Formate aus Statistik-Software gängig. Für die Metadaten existiert ein offener Standard: DDI XML.

eine effiziente Datennutzung nötig macht. Dies war für die Wissenschaft – die in diesem Bereich auch in ihrer Rolle in der Politikberatung argumentierte – ein langjähriges Desiderat.

Als Antwort hierauf wurde 2004 der Rat für Sozial- und Wirtschaftsdaten (RatSWD) konstituiert. Er entwickelte sich in der Folge zu einem starken Akteur, der sich im Dienst der forschenden Datennachfrager strategisch der Weiterentwicklung von Dateninfrastrukturen annimmt und Aushandlungsprozesse koordiniert. Die von ihm akkreditierten Forschungsdatenzentren ermöglichen durch standardisierte Zugangskonzepte den Zugriff für die Wissenschaft und auch die wissenschaftliche Nutzung der Daten.³⁵ Behörden, die ein Forschungsdatenzentrum betreiben, stellen die Daten im Rahmen ihres Auftrags kostenfrei oder mit kostendeckenden Gebühren für die Forschung bereit. Unter den Forschungsdatenzentren sind neben Behörden außeruniversitäre Forschungseinrichtungen als Datenerzeuger und Dienste-Betreiber stark vertreten, wobei sie ihrerseits mit Universitäten z.B. in Form von An-Instituten oder im Rahmen von Kooperationsverträgen über „On Campus“-Konstruktionen kooperieren. Zugriff auf sog. „Mikrodaten“ wird nur auf Antrag und nur für eine zweckgebundene wissenschaftliche Nutzung gewährt, bei besonders schützenswerten Daten wie im Falle des Forschungsdatenzentrums des Leibniz-Instituts für Bildungsverläufe nur für Angehörige berechtigter Einrichtungen. Die Forschungsdatenzentren stellen neben Datenarchiven meist auch Werkzeuge zur statistischen Auswertung bereit.

Netzwerk
akkreditierter
Forschungsdaten-
zentren

Die Nutzung der akkreditierten Forschungsdatenzentren ist beim RatSWD gut dokumentiert und erlaubt einen Einblick in die Nachfrage und den wissenschaftlichen Mehrwert. Die Nachfrage nach einer Verwendung für kommerzielle Zwecke ist nicht dokumentiert. Die Nutzung in anderen Sektoren ist bei personenbeziehbaren Daten ohnehin nur bei vollkommen anonymisierten Datensätzen (Public-User-File) möglich. Für eine Verwertung der Daten teilweise in Forschung und hauptsächlich in der Lehre werden Academic-User-Files bereitgestellt. Im RatSWD werden neben Nutzungskonzepten auch wissenschaftsgeleitete Qualitätsstandards und Policies erarbeitet, und den Datenerzeugern und Dienstbetreibern empfohlen. Daten werden in den FDZ teils aufwändig kuratiert.

Sektorabhängige
Nutzungsbedingungen

Neben den Forschungsdatenzentren gibt es weitere fachspezifische, wissenschaftsintern entwickelte Datendienste. Prototypisch für ein Datenarchiv für Long-Tail-Daten steht das *SowiDataNet* | *Datorium* von GESIS. Aus Zuwendungen sowie Gebühren zur Kostendeckung finanziert ermöglicht das Repository,

Fachspezifische,
wissenschaftsintern
entwickelte Daten-
dienste für Daten aus
Forschungsprojekten

³⁵ Stand Mai 2021: 39 Forschungsdatenzentren, ratswd.de/forschungsdaten/fdz, zuletzt abgerufen am 2.06.2021.

Daten aus der Individualforschung – rudimentär kuratiert – auf dem Wege der Veröffentlichung zu verwerten. Daneben existieren transnationale Forschungsinfrastrukturen wie CESSDA, die eine Zusammenführung von Daten bzw. Datendiensten organisieren sollen. Auch finden sich internationale Kooperationen zwischen Betreibern von Diensten, die über Kooperationen Datenzugänge im Ausland schaffen. Ein Beispiel für ein wissenschaftlich kuratiertes Datenrepositorium mit Begutachtungsverfahren und entsprechenden Reputationseffekten ist das US-amerikanische *Socioeconomic Data and Applications Center SEDAC*.

Umfangreiche Datendienste in den Teildisziplinen

Neben den Forschungsdatenzentren und fachspezifischen Repositorien finden sich in der Domäne wissenschaftsinterne Datenplattformen, die Daten, Suchmaschinen, Analysetools sowie KI-Anwendungen kombinieren. Solche umfangreichen Datendienste werden in einzelnen Teildisziplinen aufgebaut. So entwickelt ein Leibniz-Zentrum mit einem institutionellen Auftrag auf Basis der bestehenden Publikationsplattform für Psychologie das *PsychArchives* als eine Knowledge-Base einschließlich einer Data-Sharing-Plattform sowie Labor- und Analysetools für offene Daten. Datenplattformen werden auch von kommerziellen Betreibern für gewerbliche und wissenschaftliche Nutzung angeboten.

Angesichts zahlreicher kommerzieller Anbieter von Daten aus Markt- und Meinungsforschungsinstituten stellt sich die sozialwissenschaftliche Forschung zudem auf den (punktuellen) Erwerb von Daten sowie das Outsourcing von Datenanalysen – zumindest als Ergänzung dessen, was innerwissenschaftlich leistbar ist – ein.

Praxisbeispiele 5

Zu den FDZ gehört seit 2004 das Forschungsdatenzentrum der Bundesagentur für Arbeit (*FDZ BA im IAB*). Der Wissenschaft wird damit eine weltweite, adressaten- und zweckgebundene Nutzung von administrativen und Befragungsdaten zu Personen, Haushalten sowie Betrieben in Deutschland ermöglicht. Die Metadaten sind frei zugänglich, schwach anonymisierte Daten dürfen nur vor Ort, Scientific-Use-Files über Remote-Verbindung genutzt werden.

Statista ist eine weltweit tätige Plattform für statistische Daten deren Geschäftsmodell auf der passgenauen Analyse von Daten aus etwa 22.500 Quellen beruht. *Statista* richtet seine Angebote nicht nur an Nachfrager aus der Wirtschaft, sondern auch gezielt mit Campus-Lizenzen an die Wissenschaft. Mehr als 2.000 Hochschulen erwerben Campus-Lizenzen für Forschung und Lehre.

2.5 GEISTES- UND KULTURWISSENSCHAFTEN

In diese Domäne fallen Texte³⁶ und andere, primär unter dem Gesichtspunkt von Methoden eines (z.B. historisch) kontextualisierenden (Sinn-)Verstehens relevante digitale bzw. analoge Objekte.³⁷ Urheberrecht, Datenschutz wie auch Verwertungsrechte führen dazu, dass zwar eine wissenschaftliche Datennutzung gewährt werden kann, eine wissenschaftsseitige digitale „Verwertung“ bzw. Veröffentlichung dagegen nicht. Traditionell sind es Intermediäre wie Wissenschaftsverlage und Verwertungsgesellschaften (VG Wort, VG Bild-Kunst), die die Klärung der Verwertungsfragen für die geisteswissenschaftlich Forschenden regeln, sobald diese publizieren.

Verlage und Verwertungsgesellschaften als traditionelle Intermediäre

Digitale Dienste entstehen in der Domäne seit Jahrzehnten. Die in den 1960er Jahren entstandene Computerlinguistik ist an Universitäten und in der außeruniversitären Forschung gut verankert. Die digitale Editionsphilologie hat einen ähnlichen Ausbreitungsgrad. Dadurch hat sich eine beispielhafte Standard- und Methodenentwicklung und -reflexion etabliert. Auch die für die historische Forschung unverzichtbaren Historischen Grundwissenschaften wie Heraldik, Diplomatie, Kodikologie oder Epigraphik erzeugen zunehmend digitale Bestände.³⁸

Mit den Digitalen Geisteswissenschaften (Digital Humanities) hat sich in den letzten Jahren eine agile Community herausgebildet, mit Lehrstühlen (unterschiedlicher Denominationen) sowie europäisch geförderten Infrastrukturen (z.B. im Rahmen der Europäischen Roadmap-Prozesse für Forschungsinfrastrukturen/ ESFRI). Freilich sind „die“ Digital Humanities ein heterogenes Feld mit bislang noch loser Kopplung zur Mehrzahl der Methodendiskurse in den einzelnen „klassischen“ Fächern der Geisteswissenschaften und deren zahlreichen „Schulen“.

Agile Communities in digitalen Geisteswissenschaften

Datendienste werden in den Geistes- und Kulturwissenschaften in nicht voneinander abgeschotteten, aber doch distinkten Communities betrieben: Sammlungen vor allem in den Historischen Grundwissenschaften, Text-Editionen eher in Forschungseinrichtungen sowie Bilddatenbanken und Portale eher in

Projektfinanzierte und institutionell verantwortete Datendienste

³⁶ Texte – verstanden als Daten – werden vordergründig mit den Sprach- bzw. Literaturwissenschaften assoziiert. Sie stellen aber auch in allen weiteren Geisteswissenschaften (etwa der Geschichtswissenschaft und den Rechtswissenschaften sowie in den Sozialwissenschaften) eine bedeutende Datenbasis für hermeneutisch-interpretierende wie auch für quantitative bzw. algorithmische Forschungsmethoden dar. Textdateien (z.B. Kommentare, Versuchsberichte, Patentbeschreibungen) entstehen dessen ungeachtet in allen Disziplinen und sind in vielen Repositorien bzw. Datensammlungen zu finden. Digitale Forschungsmethoden für die Geistes- und Kulturwissenschaften haben daher ein hohes Potenzial, mittelfristig auch in anderen Domänen und Fachkulturen wirksam zu werden.

³⁷ Unter Bild- und Objektdaten werden Digitalisate (2D und 3D) von analogen Bildern und Objekten gefasst, sowie „born digital“ Objekte (die keine Texte sind, also etwa druckbare Architekturmodelle oder Bauteile) verstanden.

³⁸ Hohls et. al. (2016) – Historische Grundwissenschaften.

Gedächtniseinrichtungen. Dies geschieht teilweise im Rahmen des institutionellen Auftrags (Referenzkorpora), größtenteils aber in Drittmittelprojekten.³⁹ Die Datendienste richten sich an thematisch einschlägige Nutzergruppen und sind meist wenig miteinander vernetzt. Objektiv existieren in dem so entstandenen Bereich allerdings sehr gute Voraussetzungen für Interoperabilität und damit die Aggregation von Daten: In den letzten 30 Jahren wurden Standards für Texte, 2D-Daten, Metadaten, Vokabulare und Schnittstellen entwickelt.

Wissenschaftlicher
Reputationsgewinn
durch
Datenkuratierung

Die in den Geistes- und Kulturwissenschaften erzeugten Daten bzw. Referenzkorpora beziehen sich auf Lexikografie, Editionen und den Aufbau von Sammlungen. Digitale Editionen sind Beispiele, in denen die Kuratierung von Daten wissenschaftlichen Reputationsgewinn und durch die Erschließung neuer Zusammenhänge unmittelbar Erkenntnisfortschritt mit sich bringt. Dafür sind ein institutioneller Auftrag und eine etablierte Position des Anbieters von herausragender Bedeutung. Im Bereich der philologisch aufbereiteten Textkorpora schafft die editorische Leistung zusätzlich einen vermarktbareren Mehrwert. So werden einige langfristige digitale Editionsprojekte mit großen Verlagspartnern gemeinsam durchgeführt, die das Ergebnis als Produkt (z.B. als Buchreihe) vermarkten.⁴⁰

Exklusivverträge und
Vermarktung der
Daten in den
Rechtswissenschaften

Im Bereich der Rechtswissenschaften finden sich Datendienste, die den Marktwert von (Text-)Daten der öffentlichen Hand belegen. Häufig sind dies Dienste, die Texte amtlicher Werke, Gerichtsentscheidungen oder periodisch überarbeitete Rechtsnormen umfassen. Die Urteile werden von den Gerichten auf Grundlage bilateraler (aber nicht exklusiver) Verträge mit ausgewählten Dienstbetreibern publiziert. Die Auswahl orientiert sich an fachlichen Kriterien, wie der Aktualität einer Rechtsfrage, der Bestätigung eines rechtlichen Standpunktes oder der Rechtsauslegung durch unterschiedliche Gerichte in der Instanzenkette. Weitere Textgattungen sind Kommentare und Aufsätze. Da das Urheberrecht hier greift, sind Exklusivvermarktungen möglich.⁴¹ Diese Datenbestände sind meist fest in der Hand von darauf seit Jahrzehnten spezialisierten Verlagen. Ihr Dienste-Angebot richtet sich überwiegend an Nachfrager aus den juristischen Professionen, die – bzw. deren Einrichtungen (z.B. Anwaltskanzleien) – bereit sind, hohe Entgelte zu entrichten. Weitere kommerzielle Datennachfragen sind

³⁹ Hierbei ist zu unterscheiden zwischen den eine langfristige Entwicklung und ggf. den Betrieb sichernden Projekten mit einer Laufzeit von bis zu 24 Jahren und kurzfristigen, die (lediglich) prototypische Entwicklungen ermöglichen.

⁴⁰ Beispiele sind neben wissenschaftlichen Werk-Editionen auch Wörterbücher. Die Kuratierung erfolgt in drittmittelfinanzierten (Langzeit-)vorhaben zum Beispiel des Akademienprogramms oder in Forschungseinrichtungen wie etwa dem Leibniz-Institut für Deutsche Sprache.

⁴¹ Dies gilt unter Umständen auch für Gerichtsentscheidungen: Sobald Leitsätze oder Überschriften durch Zeitschriftenredaktionen oder Privatpersonen formuliert wurden, genießen sie Urheberrechtsschutz (Vgl. Dreier, Fischer et al. (Hg.) 2016 – Informationen der öffentlichen Hand, UrhG § 5 Rn. 6).

in dieser Domäne aufkommende kommerzielle Anbieter von KI-Anwendungen, die sich als rudimentäre digitale Ratgeber einsetzen lassen.⁴² Wissenschaft fungiert in diesem Teilbereich überwiegend als Datennachfrager.

In den Geschichtswissenschaften (einschließlich der Archäologie) haben sich GIS-Parameter (also Geoinformationen) als Chance für eine Erweiterung der digitalen Materialien für die historische Forschung in den letzten Jahren fest etabliert. Die historische Forschung verknüpft „multimodale“ Quellen unterschiedlichster Art, nutzt typischerweise aber auch Textdaten in großem Umfang. So bietet das Münchener Digitalisierungszentrum (MDZ) der Bayerischen Staatsbibliothek mit *DaFo* (Daten für die Forschung) kostenfreien Zugang zu Forschungszwecken zu 2,4 Millionen Druckwerken, die vielfach auch Bilder, Diagramme etc. enthalten.

Für die Forschung an historisch und kulturwissenschaftlich bedeutsamen Objekten und Artefakten sind ebenfalls umfassende internationale wissenschaftliche Datenbanken entstanden, die gemeinschaftlich betrieben werden. Exemplarisch für solche Dienste stehen das *Digital Archeological Record (tDAR)* und in Deutschland das vom DAI betriebene archäologische Datenarchiv *IANUS*, das sowohl für Daten aus der Forschung als auch aus der öffentlichen Hand (Landesarchäologien) genutzt wird.

Praxisbeispiele 6

Die vom DAI entwickelte, institutionell betriebene (und finanzierte) *IDAI.welt* aggregiert Digitalisate aus verschiedenen Einrichtungen der Altertumsforschung. Registrierte Nachfrager können eine Datenbank, Werkzeuge für Recherche und Erschließung der analogen und digitalen Bestände, 3D-Visualisierung sowie Thesauri nutzen. Teilweise liegen die Bildrechte bei anderen Einrichtungen, so dass nicht alle Objekte auf diese Weise verwertet oder gar vermarktet werden können.

Das *Bildarchiv prometheus* und das *Bildarchiv Foto Marburg* sind Beispiele für aus der Wissenschaft heraus betriebene Datendienste. *prometheus* wird von einem Verein getragen. Der Dienst integriert und erschließt („verwertet“) in einer Datenbank die (Meta-)Daten aus Gedächtniseinrichtungen (hauptsächlich Museen). Die Entwicklung erfolgt aus den Kunstwissenschaften heraus. Das *Bildarchiv Foto Marburg* wird als überregional genutzte Forschungs- und Serviceeinrichtung von der Universität Marburg betrieben und aus zentralen Haushaltsmitteln, über Drittmittelprojekte sowie Nutzungsgebühren finanziert. Eine Abfrage sowie Downloads für wissenschaftliche Nutzung sind offen, Verwertungsrechte bleiben dagegen beim Bildarchiv.

⁴² Zu nennen wären Chatbots wie Ratisbot („der erste deutsche Roboter-Anwalt“) oder Advobot („Der Legal Chatbot für Rechtsabteilungen und Kanzleien“). Die Anwendbarkeit vor allem aber die Akzeptanz und Legitimität solcher Angebote in der juristischen Praxis ist noch unklar.

Verwertung der Bestände in Gedächtnis- einrichtungen

In den Gedächtniseinrichtungen erfolgt in der Regel eine durch den institutionellen Auftrag getragene Verwertung der eigenen Bestände, wobei an den sammelnden Einrichtungen selbst die Erschließung der Bestände nicht immer systematisch, sondern in Form inkrementeller Sammlungen über konsekutive Drittmittelprojekte erfolgt.⁴³ Die erschlossenen Daten werden – so der Eindruck – bislang seitens der kooperierenden Wissenschaft nicht systematisch in übergreifende Katalogsysteme eingebracht, so wie es beispielsweise im Bereich der Long-Tail-Daten Praxis ist (vgl. 2.6). Die Gedächtniseinrichtungen nutzen ihre Internetportale zur Verwertung von Faksimiles, Bildern und Objekten. Wie die dahinterliegenden Daten tatsächlich gesichert werden, kann (noch) nicht abschließend beantwortet werden. Die Strategien scheinen diesbezüglich unterschiedlich zu sein. Allerdings bilden sich zwischen den Gedächtniseinrichtungen Verbundstrukturen, in denen übergreifende Portale und Nachweisysteme betrieben werden. Im Fall der Erschließung analoger Werke durch kulturelle Einrichtungen können einzelne Institutionen, regionale Verbände und überregionale sowie internationale Akteure Datenerzeuger bzw. Betreiber der Angebote sein.

Ausschließlichkeits- vereinbarungen in Digitalisierungs- vorhaben

Auch kommerzielle Anbieter aggregieren Daten aus den Gedächtniseinrichtungen bzw. führen in deren Auftrag Digitalisierungsprojekte durch und vermarkten die so gewonnenen Daten. Anders als bei den Forschungsdaten erlaubt das Datennutzungsgesetz bei Gedächtniseinrichtungen Ausschließlichkeitsvereinbarungen mit einer Laufzeit von bis zu zehn Jahren. Für den vereinbarten Zeitraum ist ein kostenloser Zugriff auf hochwertige Daten für die Wissenschaft nur begrenzt möglich. Aufgrund der Unterschiede in nationalen Regelungen des Urheberrechts können internationale Diensteanbieter offene Daten kommerzialisieren; was im Fall der kommerziellen *Bildagentur Getty* zu einer äußerst umstrittenen Aneignung geführt hat.⁴⁴

Zivilgesellschaftliche und gewerbliche Verwendung der Daten

Für die in den Gedächtniseinrichtungen und Wissensspeichern (wie Museen, Bibliotheken, Akademien und Archive) sowie in der Forschung erzeugten Daten sind neben der Wissenschaft der Bildungsbereich und die Zivilgesellschaft die wichtigsten Datennachfrager. Eine gewerbliche Verwendung der Daten aus der Domäne der Geistes- und Kulturwissenschaften ist auch gängig. Digital Artefakte haben hier einen teils hohen Marktwert. So werden Digitalisate von Kunstwerken oder archäologischen Artefakten zu Werbezwecken eingesetzt (etwa für regionalen Tourismus), bringen Tantiemen aus dem Kulturbereich ein oder dienen als Vorbilder für virtuelle Welten in der Computerspiel-Industrie. Lexikalische Daten werden für die Entwicklung von Spracherkennungs-Software verwendet

⁴³ Klaffki et al. (2018) – Stand der Kulturgutdigitalisierung in Deutschland.

⁴⁴ United States District Court (2016) – *Highsmith v. Getty Images*.

oder als Trainingsdaten für KI-Anwendungen. Neben der klassischen hermeneutisch-interpretierenden Nutzung besteht an den digitalen Daten somit auch ein Interesse ganz anderer, eher technisch oder marktorientierter Nutzergruppen.

2.6 LONG-TAIL-DATEN AUS DER FORSCHUNG

Eine Reihe der für dieses Positionspapier analysierten Datendienste unterstützt die Publikation und Archivierung von sog. „Long-Tail-Daten“ aus der Individual- und Projektforschung. Die betreffenden Datensätze sind in der Regel abgeschlossen und von Aufbau, Umfang und Format her heterogen. Sie werden – oftmals im Zusammenhang mit der Ergebnispublikation – als eigenständige digitale Objekte publiziert und von den Datendiensten in der Regel auch archiviert (mindestens für einen verabredeten Zeitraum, etwa zehn Jahre).

Heterogene Daten
aus der Individual-
und Projektforschung

Einige dieser Dienste sind disziplinär ausgerichtet, d.h. zielen auf eine Fachgemeinschaft/Community und werden in der Regel von einer Forschungseinrichtung getragen. Darüber hinaus hat sich seit etwa zehn Jahren eine ihre Dienste teilweise weltweit anbietende Landschaft von Diensten entwickelt, die sich generell an Forschende und ihre Einrichtungen richten, also fachübergreifend bzw. fachlich „agnostisch“ sind. Sie werden im Folgenden als „generische Datenrepositorien“ bezeichnet. Betreiber sind sowohl in der Wissenschaft selbst wie unter Verlagen zu finden.

Generische
Datenrepositorien
diverser Anbieter

Eine Vielzahl generischer Datendienste unterstützen mindestens die Verwertung von Daten (Weitergabe an eine Öffentlichkeit). Einige publizieren aber auch digitale Objekte aller Art, wie graue Literatur oder Code. Es handelt sich also nicht in jedem Fall um reine „Datenarchive“. Überhaupt spielt die „Wissenschaftlichkeit“ der (Nach-)Nutzung keine prägende Rolle. Vielmehr finden sich im Bereich der Dienste Ausprägungen, die Ähnlichkeiten zu Plattformdiensten aufweisen: Nutzerinnen und Nutzer können private Speicherkapazitäten hinzubuchen bzw. hinzukaufen, und einige Anbieter verfügen über recht weitreichende Integrationen mit den Publikationsplattformen von Verlagen, was eine Bereitstellung der Daten parallel zum Einreichungs- und Begutachtungsprozess für die zugehörigen Artikel erleichtert. Anders als bei wissenschaftsgeleitet kuratierten Datenbeständen findet seitens der Betreiber in der Regel nur eine cursorische Auswahl und Qualitätssicherung statt. So wird primär darauf geachtet, dass die hochgeladenen Datensätze keine personenbeziehbaren Daten enthalten und mit einem Minimum an Metadaten sowie im Falle einer Veröffentlichung mit einer Lizenz versehen sind.

Fokus auf Bedarfe
der Datenerzeuger

Die Wissenschaft verhält sich mit zahlreichen eigenen Repositorien durchaus ein Stück weit als „Selbstversorger“. In Deutschland findet sich seit ca. 2015 eine wachsende Zahl von Inhouse-Angeboten forschender Einrichtungen, die sich als Service an die eigenen Mitarbeiter richten und überdies der

Inhouse-Angebote
forschender
Einrichtungen

Kommunikation eigener Forschungsleistungen dienen. Meist mit Hilfe von Drittmitteln aufgebaut, werden diese lokalen Dienste üblicherweise aus Haushaltsmitteln der Institution im Rahmen ihres Auftrags betrieben.⁴⁵ Von der technischen Seite aus betrachtet werden ganz unterschiedliche Lösungen gewählt. Einige lokal betriebene Dienste erscheinen als Erweiterungen schon bestehender Open Access-Repositoryn, die – eigentlich zur Zweitverwertung wissenschaftlicher Literatur aufgesetzt – für weitere digitale Objekte geöffnet werden. Andere Einrichtungen nutzen spezialisierte Software für das Datenmanagement.⁴⁶ Zudem existieren technische Dienstleistungsangebote, wie das Hosting institutioneller Instanzen bei *Figshare* oder bei der wissenschaftsinternen Alternative *RADAR* vom FIZ Karlsruhe.

Praxisbeispiele 7

Das Unternehmen *Figshare* betreibt ein Datenrepository mit guten Workflow-Integrationen zu den Publikationsplattformen internationaler Verlage. Es positioniert sich zudem im Segment der institutionellen Repositoryn als Anbieter von Software-as-a-Service bzw. Infrastructure-as-a-Service, d.h. es bietet Forschungseinrichtungen eigene Instanzen an, die sie entweder komplett selbst betreiben oder durch *Figshare* hosten lassen können.⁴⁷ Die technischen Dienstleistungen von *Figshare* werden gut angenommen, vor allem in den USA, Großbritannien, aber auch in den Niederlanden.⁴⁸

Das Repository von *Dryad*, einer wissenschaftsnahen Non-Profit Initiative mit Sitz in den USA, ist 2008 auf Impuls einiger Zeitschriftenverlage als kooperativ organisierter Dienst entstanden. Es wird mittlerweile als Public-Private-Partnership betrieben über einen Verein, in dem überwiegend wissenschaftliche Organisationen, aber auch einige Zeitschriftenverlage Mitglieder sind. Es verfügt über gute Workflow-Integrationen. *Dryad* positioniert sich ebenfalls als Kooperationspartner und technischer Dienstleister. Die California Digital Library und der vom CERN betriebene Dienst *Zenodo*, der vor allem den EU-Markt bedient, sind 2018 und 2019 Kooperationen mit *Dryad* eingegangen, um entsprechende Funktionen weiterzuentwickeln.

Das allein die Wissenschaft adressierende Repository *RADAR* wird seit 2017 (nach einer von der DFG geförderten Aufbauphase) vom FIZ Karlsruhe betrieben und weiterentwickelt. Nach Vertragsabschluss mit institutionellen Kunden bzw. Projekten werden leistungsabhängig gestaffelte Nutzungsgebühren für Uploads erhoben.

⁴⁵ Eine Studie aus der Schweiz weist auf den teils prekären Status solcher Dienste hin, die gelegentlich auch wieder eingestellt werden. Heyde (2019) – Open Research Data Landscape Switzerland.

⁴⁶ Open Source-Software, hervorgegangen aus der Harvard Universität; dataverse.org, zuletzt abgerufen am 2.06.2021.

⁴⁷ Nach eigenen Angaben ist dies das eigentliche Geschäftsmodell, das es erlaubt, die kostenfreien Angebote der Datenpublikation auf der zentralen Figshare-Instanz mitzufinanzieren. Vgl. Frank (2019) – A look into Figshare.

⁴⁸ In re3data.org finden sich neben Figshare 33 weitere institutionelle Instanzen von Universitäten und weiteren Forschungseinrichtungen.

Zusätzlich zum Basis Service (Archivierung und Veröffentlichung von Forschungsdaten, Schnittstelle für Peer-Review, Metriken) sind erweiterte vertragliche Leistungen wie Implementierung einer Instanz an der Einrichtung oder die Anpassung an das Corporate Design buchbar. Verträge mit individuellen Forscherinnen und Forschern sind nicht vorgesehen. Hervorgehoben wird die Umsetzung der FAIR-Prinzipien.

Insgesamt kann konstatiert werden, dass der Markt für Datenpublikationsdienste noch stark in Bewegung ist. Die Dienste bedienen einen Bedarf nach unkomplizierter Veröffentlichung von digitalen Objekten mit starken Parallelen zum etablierten Publikationswesen für wissenschaftliche Literatur.⁴⁹

Unter den Funktionalitäten erscheint die Integration mit Publikationsworkflows von Zeitschriftenverlagen als Vorteil im Wettbewerb. Der Zugang erfolgt oftmals auf dem Wege der Lektüre der zugehörigen Artikel, oder auf dem Wege der persönlichen Kommunikation (Peer-to-Peer Teilen von Daten).⁵⁰ Auf andere Formen der Datennachfrage, wie die freie Suche nach Datensätzen in den Repositorien selbst, sind die Dienste weniger gut eingerichtet.⁵¹ Suchparameter und Filter sind oftmals auf wenige Auswahlmöglichkeiten begrenzt. Das liegt auch an der Erschließung der Datensätze, bei der gerade die generischen Datenrepositorien vor besonderen Herausforderungen stehen: hierzu zählt die Heterogenität der Datensätze, aber auch der meist begrenzte Ressourceneinsatz für die Kuratierung der Metadaten.⁵²

Integration mit
Publikationsworkflows
von Zeitschriftenverlagen

⁴⁹ Der Wissenschaftsrat warnt, dass eine „nachhaltige Sicherung der Daten über lange Zeiträume [...] nicht bspw. durch die Auslagerung an Repositorien von Verlagen garantiert werden“ kann. WR (2020) – Zum Wandel in den Wissenschaften, S. 26.

⁵⁰ Einige der Dienste bieten zum Beispiel auch für nichtveröffentlichte (im Privatmodus hochgeladene) Datensätze weitergabefähige Links an.

⁵¹ Vgl. dazu auch: Assante et al. (2016) – Scientific Data Repositories.

⁵² Der Rfll behält sich vor, Repositorien und ihre Rolle in der Nutzung und Verwertung von Forschungsdaten an anderer Stelle vertieft zu betrachten und hierzu Empfehlungen auszusprechen.

3 ÜBERGREIFENDE MERKMALE UND AUSPRÄGUNGEN VON DATENDIENSTEN

Im Rahmen seiner Fallbeispielanalyse hat sich der Rfll auf diejenigen Merkmale und Ausprägungen von wissenschaftlich relevanten Datendiensten konzentriert, die einen Dienst attraktiv für die jeweilige(n) Zielgruppe(n) machen, die (aus wissenschaftlicher Sicht) Mehrwerte mit sich bringen oder aber unerwünschte Nebeneffekte haben. Dabei wurden fünf übergreifende Aspekte berücksichtigt:

Vergleichsdimensionen

Die am weitesten zweckübergreifende Vergleichsdimension ist der **Umfang und die jeweils eingesetzte Technologie von Angeboten** eines Dienstes. Die Vermarktung digitaler Daten erfolgt z.B. zunehmend über assoziierte Services oder Tools, was unter dem Gesichtspunkt der wissenschaftlichen Nutzbarkeit Folgen hat. So liegt der Gedanke nahe, dass Dienste, die ihren technologischen Einsatz auf domänenübergreifende Plattformangebote konzentrieren, für die Wissenschaft qualitativ eine eingeschränktere Nutzbarkeit bieten als Dienste, die Schnittstellen und digitale Analysewerkzeuge für eine Nutzung entwickeln, die an die Forschungsfragestellungen konkreter wissenschaftlicher Forschungsfelder rückgekoppelt oder sogar aus diesen heraus entwickelt worden sind.

Aus Sicht der Wissenschaft sind außerdem **Akteurskonstellationen** hervorzuheben, die Aufschluss darüber geben, welche Stakeholder eigentlich ein Interesse an der funktionalen Ausgestaltung und Arbeitsweise des jeweiligen Dienstes haben. Es geht also um Rollenverteilungen und Partizipation in den Organisationsmodellen. Aus welchem sozialen System und aus welchem organisatorischen Feld heraus das jeweilige Angebot entwickelt wird, ist ein Faktor, der wissenschaftsförderliche oder gute wissenschaftliche Praxis eher beeinträchtigende Geschäftsmodelle in bedeutendem Maße bestimmen kann. Akteurskonstellationen spiegeln sich des Weiteren auf der konkreten Ebene von Organisation und Betriebsmodell in der **Governance** eines Dienstes wider, verstanden als die operative Entscheidungsebene, die die Auswahl und Kuratierung von Inhalten steuert sowie über den Einschluss/Ausschluss bestimmter Akteure oder Nutzergruppen entscheidet. Letzteres führt zu der Frage, durch welches **Zugangsregime** ein Dienst gekennzeichnet ist und durch welches Geschäftsmodell er seine **Finanzierung** sichert. Auch hierzu wurden Ausprägungen vergleichend analysiert.

Die folgende Tabelle zeigt im Sinne eines Überblicks, wie vielschichtig die übergreifenden Merkmale und Ausprägungen von Diensten ausfallen. Diese wurden der Analyse der Fallbeispiele nicht im Sinne eines Suchrasters übergestülpt, sondern vielmehr induktiv aus dem empirischen Vergleich einzelner Dienste gewonnen.⁵³

Table 2: Übergreifende Merkmale von Datendiensten und ihre Ausprägungen

Umfang, Technologie und Inhalte der Dienste
Datenarchive
Datenplattformen
Wissensdatenbanken bzw. „Knowledge Bases“
Akteure und ihre Rollen
Datensubjekte
Datenerzeuger
Datennachfrager
Geldgeber
Anbieter/Betreiber
Technische Dienstleister
Vertriebspartner und „Broker“
Kuratoren
Governance – Entscheidungsprozesse und Verantwortung
Forum- oder Marktplatzmodell
Redaktions- oder Editionsmodell
Zugangsregimes
Zugriff nach Registrierung
Zugriff auf Antrag
Offener Zugang zu Inhalten
Finanzierung und Monetarisierung
Zuwendungen aus institutioneller Finanzierung und Projektmitteln der Forschungsförderung
Community Support: Mitgliedsbeiträge und Spenden
Monetarisierung: Gebühren, „Freemium“ und Investmentkapital

3.1 UMFANG, TECHNOLOGIEN UND INHALTE DER DIENSTE

Für die Datennutzung in der Forschung sind insbesondere Plattformtechnologien interessant, sofern hier Daten mit Werkzeugen bereits verbunden sind bzw. unaufwendig verbunden werden können. Im Rahmen der Fallbeispielanalyse wurden nur Plattformdienste betrachtet, die Inhalte verwerten oder vermarkten. Darüber hinaus entwickelt sich inzwischen ein Markt für digitale Plattformen, die zum Beispiel Laborgeräte oder komplexere Analysetools einbinden und somit direkt den Forschungsprozess (vor einer Veröffentlichung) substantiell unterstützen.

Verbindung von Daten und Werkzeugen durch Plattformdienste

⁵³ Wie in Kapitel 1 bereits ausgeführt, erhebt die vom Rfll durchgeführte Fallbeispielanalyse keinen Anspruch auf Repräsentativität und basiert auf Informationen soweit diese im Rahmen einer Webrecherche und Analyse öffentlich zugänglicher Dokumente erhoben werden konnten.

Zunehmend wird auch das Potenzial deutlich, das in der (fachlich) durchdacht aufgebauten Kombination von wissenschaftlicher Literatur und Daten liegt („Wissensdatenbanken“). Literatur ist weltweit aufgrund langer bibliothekarischer Traditionen sehr gut über Metadaten und Schnittstellen erschlossen. Die Entwicklung im Bereich der Datenrepositorien schließt hier an. Sie scheint zum einen die Evolution von kombinierten Informationsangeboten zu befördern (Artikel plus Datensatz, „Enhanced Publications“). Zum zweiten bilden sich spezialisierte Dienste heraus, die Daten für bestimmte Forschungszwecke aggregieren, teils verbunden mit erheblichem Kuratierungsaufwand.⁵⁴

Datenarchive – Datenplattformen – Wissensdatenbanken

Unter **Datenarchiven** werden Services für Datenspeicherung, Veröffentlichung und Auffindbarkeit gefasst.⁵⁵ In den meisten Fällen erschließen die Nutzer ihre Daten mit Metadaten selbst, die vergebenen Lizenzen für die einzelnen digitalen Objekte können entsprechend heterogen sein (vgl. dazu auch 2.6). Die grundlegenden Umfänge der Dienstleistungen erscheinen weitgehend standardisiert. Einige Wettbewerber bieten Sonderleistungen an,⁵⁶ für die teilweise gesonderte Gebühren erhoben werden.⁵⁷

Auf **Datenplattformen** werden Daten in Verbindung mit Analyse- und Visualisierungstools bzw. mit Infrastruktur wie beispielsweise Rechen- und Speicherkapazitäten verwertet bzw. vermarktet. Nutzerinnen und Nutzer sollen die Plattformen als ein Werkzeug für den (eigenen) Datenverarbeitungsprozess verwenden können. Datenplattformen ähneln daher den virtuellen Arbeitsumgebungen („Virtual Research Environments“, VRE)⁵⁸, die eine Zeitlang auch durch öffentliche Mittel

⁵⁴ Der Wissenschaftsrat sieht die Aggregation von Daten aus unterschiedlichsten Quellen als eine Möglichkeit, die in ihrer Generierung besonders kostenintensiven Daten über sektorale Grenzen hinweg zu nutzen, weist aber auf die dabei entstehenden ethischen Herausforderungen im Falle von personen- oder gruppenbezogenen Daten hin. WR (2020) – Zum Wandel in den Wissenschaften, S. 17. Diese ethischen Herausforderungen müssen im Rahmen der Kuratierung besonders berücksichtigt werden.

⁵⁵ Auch als Archival Data Repositories bezeichnet, vgl. Anderson et al. (2017) – Support of Core Data Resources for the Life Sciences, S. 3.

⁵⁶ Als Standardleistungen werden Archivierung, Erfassung der technischen und administrativen Metadaten sowie Veröffentlichung angeboten. Der Leistungsumfang beinhaltet oft auch die Bereitstellung maschinenlesbarer Schnittstellen und Metriken, seltener sind Versionierung, Rechtemanagement, Einbindung in Reporting-Systeme der Förderer, Integration mit Journals bzw. Verlagsplattformen und weiteren Systemen wie beispielweise Laborbücher, Kuratierung sowie Use Files (scientific, reviewer).

⁵⁷ Generische und fachspezifische Datenarchive werden zunehmend auch zur Veröffentlichung von Preprints genutzt. Zu dieser Form des Publizierens wird der Rfll zu einem späteren Zeitpunkt Stellung nehmen.

⁵⁸ Eine virtuelle Forschungsumgebung „[...] unterstützt potentiell den gesamten Forschungsprozess – von der Erhebung, der Diskussion und weiteren Bearbeitung der Daten bis zur Publikation der Ergebnisse – während sie technologisch vor allem auf Softwarediensten und Kommunikationsnetzwerken basiert.“ Sie gewährt „[...] Zugang zu realen Forschungsressourcen, wie z.B. Daten, Analyse- und Bearbeitungswerkzeuge, wissenschaftliche Geräte und Dienste.“ Arbeitsgruppe Virtuelle Forschungsumgebungen in der Allianz der Wissenschaftsorganisationen (2011) – Definition Virtuelle Forschungsumgebung. Bis heute existieren: FuD – Forschungsnetzwerk und Datenbanksystem (Uni Trier) und TextGrid (als ein e.V.).

stark gefördert wurden. Auf den Plattformen kann ein Datennachfrager zwischen verschiedenen Applikationen wählen und erhält Zugang zu Rechen- und Speicherkapazitäten. Dies macht ihn unabhängig von der technischen Ausstattung der Einrichtung, in der er tätig ist. Die Plattformen bieten zudem Entwicklern von Applikationen die Möglichkeit zur Verwertung/Vermarktung von Software.

Wissensdatenbanken („Knowledge Bases“) aggregieren Daten aus unterschiedlichen Quellen (Literatur, Bibliografien, Patente, Datenarchive). Ihr Hauptangebot an die Wissenschaftler ist das Suchen und Finden. Ihre Inhalte sind dynamisch, d.h. werden laufend aktualisiert und ergänzt. Je nach eingesetzter Technologie sind die Übergänge zwischen solchen Informationsinfrastrukturen und den Datenplattformen fließend. Die Erschließung von Inhalten („Content“) erfolgt teils durch manuelle Kuratierung, teils durch Einsatz von KI-Technologien. Die vom RfII betrachteten Dienste erschließen zum Beispiel Literatur, Patente und Daten für die Lebenswissenschaften bzw. die Chemie. Sie weisen einige Parallelen zu den systematisch aufgebauten Fachinformationsdiensten (FID) in Deutschland auf. Diese gingen aus den Sondersammelgebieten der wissenschaftlichen Bibliotheken hervor, durch die arbeitsteilig Spezialliteratur für die beteiligten Fächer beschafft wurde. Mit der Umstellung auf das Konzept des „Fachinformationsdienstes“ werden seit 2014 vermehrt digitale Ressourcen erworben und überregional angeboten sowie größere Datenmengen unterschiedlicher Daten akquiriert und für spezifische Zielgruppen erschlossen. Akquise und Erschließung werden an den Bedarfen der Forschung ausgerichtet.

Aggregierte Daten
verbunden mit
Wissensbeständen
eines Faches

Praxisbeispiele 8

Die Objektdatenbank *ARACHNE* basiert auf analogen Beständen. *ARACHNE* aggregiert Daten (Metadaten und Images) aus Beständen von Forschungs- und Gedächtniseinrichtungen, die systematisch oder in Forschungsprojekten entstehen. Die Datenbank wird unter dem Service iDAI.world mit weiterem Content (iDAI bookbrowser – Integration von Dokumenten in die Objektstruktur) und Tools (Thesauri, Erschließungswerkzeuge) verbunden. Hiermit vergleichbar sind unter anderem Datenbanken wie das *Deutsche Textarchiv* oder *DeReKo*, die zur Analyse bereitgestellte Daten teilweise mit eigenen, teilweise mit externen Tools verknüpfen.

Die Wissensdatenbank *TAIR* ist auf die speziellen Bedürfnisse einer kleinen Forschungscommunity ausgerichtet, die an *Arabidopsis thaliana* forscht, einem wichtigen Modellorganismus in der molekularen Pflanzenphysiologie. Das Produkt ist ein „annotiertes Genom“, der Dienst verzeichnet also Funktionen und Varianten für die bekannten Gene, die aus veröffentlichten Fachartikeln und Sequenzdaten zusammengestellt werden. Ähnliche „Knowledge Bases“ finden sich in den Lebenswissenschaften auch für zahlreiche andere Modellorganismen.

3.2 AKTEURE UND IHRE ROLLEN IM INFRASTRUKTURKONTEXT

Viele der betrachteten Dienste werden nicht durch einen Akteur allein, sondern in Partnerschaft mit anderen gestaltet – oftmals über die Wissenschaft hinaus mit Stakeholdern bzw. Partnern aus Wirtschaft, öffentlicher Hand oder der Zivilgesellschaft.⁵⁹

Ausdifferenzierte
Arbeitsteilung
zwischen Akteuren
aus der Forschung
und von
Intermediären

Die jeweiligen Konstellationen von Akteuren in bzw. hinter den Diensten können sowohl bezüglich der Arbeitsteiligkeit wie auch der sektoralen Zugehörigkeit überaus vielfältig sein. Die auf den ersten Blick bilaterale Beziehung zwischen „Anbietern“ von Daten und den „Nutzern“ erweckt den Eindruck einer kausalen Gerichtetheit, die real allerdings kaum vorkommt. Ein Datendienst kann beispielsweise Angebote beinhalten, in denen Erzeuger von Forschungsdaten sich eines Intermediärs bedienen, um die Daten zu prozessieren, und eines anderen, um diese zu verwerten oder sogar zu vermarkten. Gegebenenfalls wird ein Dienst kollaborativ durch eine Gruppe von Akteuren angeboten, in der die Partner arbeitsteilig Aufgaben übernehmen.

Zu unterscheiden sind mindestens acht mögliche Rollen, in denen Akteure die Ausgestaltung (Governance) eines Dienstes prägen können. Neben Datenerzeuger und Datennachfrager sind dies u.a. Geldgeber sowie Betreiber/Anbieter eines Dienstes. Gelegentlich finden sich darüber hinaus weitere intermediäre Rollen, wie Vertriebspartner oder Broker, technische Dienstleister sowie Kuratoren. Hinzu kommen in einigen Domänen die beforschten Personen (Datensubjekte). Die Aufschlüsselung der jeweiligen Rollen ermöglicht eine differenzierende Beschreibung der vorgefundenen Konstellationen in den Fallbeispielen. Das jeweilige Rollenmodell wird im Folgenden kurz vorgestellt und mit Beispielen unterlegt.

3.2.1 Datensubjekte – Datenerzeuger – Datennachfrager

In einigen Domänen sind Erhebungen von Daten natürlicher Personen Grundlage der Forschungsarbeit: von Patienten, Probandinnen oder Klienten in der Medizin über Teilnehmerinnen an Befragungen, Studien, experimentellen Beteiligungsformen aller Art bis zu „Forschungsadressatinnen“ und Forschungspartnern in der Ethnographie.

Vereinbarungen mit
beforschten oder
datengebenden
Personen

Für die Nutzung und Verwertung von Forschungsdaten ist dieser Personenkreis insofern relevant, als Vereinbarungen mit der beforschten oder datengebenden Person zu treffen sind und Einwilligungserklärungen sowie weitere Dokumente

⁵⁹ Die DFG wies auf die Bedeutung der strategischen Zusammenarbeit mit der Industrieforschung bezüglich der Entwicklung digitaler Technologien hin. DFG (2020) – Digitaler Wandel in den Wissenschaften, S. 13.

möglicherweise über einige Zeit auch gemanagt werden müssen. Auch bei Zusage der faktischen oder gar der absoluten Anonymisierung sind nicht wenige Probanden zurückhaltend bei der Erteilung einer Genehmigung zur Weitergabe der Daten. Auch auf nachträgliche Datenlöschungen können Versuchspersonen unter Umständen einen Anspruch haben. Unabhängig davon ist es nicht unüblich, dass Probanden Forschungsergebnisse (bzw. Teile davon) mitgeteilt werden und dass eine finanzielle Entschädigung für den Aufwand gezahlt wird (gerade bei Längsschnittstudien mit wiederholten Befragungen). Dies ist auch ein Kostenfaktor für die Datenerhebung, der ggf. auf die Bereitstellung durchschlägt.

Die Beziehung zur beforschten Person kann somit für die Gestaltung eines Zugangsregimes zu Forschungsdaten, die Governance und die Finanzierung eines Dienstes mitbestimmend sein. Natürliche Personen werden in Diskursen über Verbraucherdaten manchmal als „Datenproduzenten“ bezeichnet. Die Rolle erhält im hier betrachteten Forschungskontext jedoch in Anlehnung an die Praxis im Bereich des Datenschutzes die Bezeichnung **Datensubjekt**.⁶⁰

Mitbestimmung bei Daten(nach)nutzung

Im Forschungsprozess werden Daten zum Beispiel durch Messungen, Erhebungen und Auswertungen erzeugt bzw. „gewonnen“. **Datenerzeuger** können einzelne Forscherinnen und Forscher oder Forschungsgruppen sein, die Daten im Kontext einer wissenschaftlichen Publikation veröffentlichen bzw. andere nachnutzbare Datenprodukte hervorbringen.

Datenerzeugung führt oftmals zu einem erst inkrementellen, später dann gezielten und koordinierten Auf- und Ausbau von Datenbeständen in der Einzel- und Verbundforschung. Die Datenbank *ChemSpider* ist ein Beispiel für langjährige Eigeninitiative einzelner Wissenschaftler, die manchmal mit dem Ziel entstehen, ein Datenkonvolut der Fachgemeinschaft zur Nachnutzung zur Verfügung zu stellen. Manchmal werden sie zunächst als private „Arbeitshilfe“ begründet, die später – weil nachweislich (auch wenn zunächst „nur“ kollegial) gefragt – veröffentlicht und „institutionalisiert“ wird.⁶¹

⁶⁰ Definiert als „eine natürliche Person, auf die sich die Daten beziehen“ ([dz-cp.de/dateien/pdf/poc/einzelbeitraege/poc-2-2018-beileger-leitfaden-eu-dsgvo](https://www.dz-cp.de/dateien/pdf/poc/einzelbeitraege/poc-2-2018-beileger-leitfaden-eu-dsgvo)). In der Datenschutzgrundverordnung (DSGVO) wird der Begriff nicht verwendet, sondern von einer „natürlichen Person“ oder einer „betroffenen Person“ gesprochen. In Verbindung mit der DSGVO hat sich im öffentlichen Diskurs die Bezeichnung „Betroffener“ etabliert. So werden die Bezeichnungen „Betroffene(r)“, „natürliche Person“, „Datensubjekt“ und „betroffene Person“ zurzeit als Synonyme benutzt ([dsgvo-gesetz.de/art-4-dsgvo/](https://www.dsgvo-gesetz.de/art-4-dsgvo/)), zuletzt abgerufen am 2.06.2021.

⁶¹ So berichtet Antony Williams, der Entwickler von ChemSpider von seinem „Hobbyprojekt“: „What originally started as a hobby project to give back something to the chemistry community has become one of the primary internet resources for Chemistry. And this from home built computers in a basement, with no funding and a team of volunteers.“ Die Sammlung wurde 2009 von der Royal Chemistry Society übernommen. Vgl. RSC (2009) – RSC acquires ChemSpider.

Datenerzeugung durch Digitalisierung analoger Sammlungen

Digitale Daten entstehen nicht nur durch Datenerhebung im engeren Wortsinn, sondern auch durch die Digitalisierung analoger Sammlungen. Insofern sind nicht nur Forschungseinrichtungen, sondern auch Gedächtniseinrichtungen und Universitätsbibliotheken als Datenerzeuger zu betrachten. Auch klassische Rechen-, Schreib- und Publikationsvorgänge sowie das Generieren von Software sind Datenerzeugung, insbesondere (obgleich nicht erst), wenn das Ergebnis in digitaler Form festgehalten wird.⁶²

Wissenschaft als Datennachfragerin für sektorübergreifende Daten

Viele Dienste haben **Datennachfrager** aus verschiedenen Sektoren. Innerhalb der Wissenschaft ist die Nachfrage nach Forschungsdaten unterschiedlich stark ausgeprägt. So können Replikationsstudien, die Nutzung von Referenzkorpora oder von Daten aus Längsschnittstudien als Beispiele für regelhafte Nachnutzung („Science-to-Science“) betrachtet werden. Wissenschaft tritt aber auch als Datennachfragerin an nichtwissenschaftliche Akteure heran, etwa wenn es um Daten aus der öffentlichen Verwaltung (Statistische Ämter, Geodaten), aus publizistischen Medien oder aus Gedächtniseinrichtungen und natürlich auch aus Unternehmen geht. Wissenschaftliche Datennachfrage in Richtung der öffentlichen Verwaltung kommt nicht nur aus der Forschung, sondern auch aus der Lehre. Hierfür entwickeln Anbieter teils eigene „Produkte“, wie z.B. die Campus Files der statistischen Ämter oder des Forschungsdatenzentrums der Bundesagentur für Arbeit.

Erdbeobachtungsdaten aus den internationalen Satellitenprogrammen sind ein Beispiel dafür, wie offene Daten von verschiedenen Intermediären für die Entwicklung jeweils zielgruppenspezifischer Dienste genutzt werden (verschiedene Wirtschaftszweige, Forschung). So bereitet für die Datennachfrage aus der Wissenschaft (insbesondere, wenn es um historische Satellitenbilder benötigende Forschungsprojekte geht), aus der Zivilgesellschaft und aus der Wirtschaft ein kommerzieller Anbieter (*Google Earth Engine*) die Daten mit Blick auf die verschiedenen Bedarfe auch unterschiedlich auf.

Unterschiedliche Zugangsregimes für Wirtschaft und für Wissenschaft

In Domänen mit einem hohen Anteil industrieller Forschung, etwa der Chemie, stammen Nachfrager gleichermaßen aus Wirtschaft und Wissenschaft, wobei die Zugangsregimes für beide Gruppen unterschiedlich sein können. In einigen der untersuchten Fallbeispiele – i.d.R. bei innerwissenschaftlichen Diensten – handelt es sich bei Datennachfragern und Datenerzeugern um die gleichen

⁶² Auch analoge Forschungsdaten sind „Daten“, auch wenn aus Sicht des Forschungsdatenmanagements die Digitalisierung und digitale Daten im Vordergrund des Interesses stehen. In diesem Sinne hat der RfII einen weiten Begriff von „Forschungsdaten“ vorgeschlagen und geprägt. Vgl. RfII (2016) – Leistung aus Vielfalt, S. A-13.

Akteure, die sich gemeinsame Ressourcen für ihre Arbeit aufbauen;⁶³ die Rollenmuster sind also in den empirisch auffindbaren Governance-Strukturen der wissenschaftlichen Dienste nicht immer trennscharf.

3.2.2 Geldgeber – Anbieter – technische Dienstleister – Vermittler

Wo ein Angebot sich nicht allein über Einnahmen/Gebühren refinanziert (was in der öffentlichen Forschung selten der Fall ist), sichern **Geldgeber** (bzw. dort, wo es sich um staatliche Zuwendungen handelt: „Förderer“) sowohl die Entwicklung als auch den nachhaltigen Betrieb von Diensten. Fördermittel werden von unterschiedlicher Seite gerade auch für einen Aufbau bzw. eine Erweiterung bestehender Datenangebote gewährt.

Kaum Refinanzierungschancen des Betriebs wissenschaftseigener Datendienste über Gebühren

Bei forschungsbezogenen Datendiensten handelt es sich überwiegend um Geldgeber der öffentlichen Hand – primär um Organisationen der Forschungsförderung, im Einzelfall auch um Bundesbehörden oder transnationale staatliche Netzwerke. Hinzu kommt die institutionelle Finanzierung von Datendiensten im Rahmen von Forschungseinrichtungen. Voraussetzung ist hier, dass der Betrieb des Dienstes Teil des institutionellen Auftrags der Einrichtung ist, er muss also mit den Trägern vereinbart sein. Punktuell fungieren auch Akteure aus Wirtschaft und Zivilgesellschaft als Geldgeber: zum Beispiel in Public-Private-Partnerships, als Investoren von Start-ups oder sogar als Spender.

Überwiegende Finanzierung durch die öffentliche Hand

Die Rolle des **Anbieters** oder **Betreibers** eines Dienstes umfasst den organisatorischen und geschäftlichen Betrieb eines Dienstes. Diese Rolle kann auch in Kombination mit der des Datenerzeugers vorkommen, etwa dort wo eine Einrichtung ihre Daten über einen eigenen Dienst für die Nachnutzung zur Verfügung stellt und ggf. „verwertet“ („Vermarktung“ ist dies allerdings nur dann, wenn es sich um ein kommerzielles Unternehmen handelt, s. Kap. 1.2).

Anbieterrolle

Es finden sich auch zahlreiche Betreiber, die Inhalte anderer Datenerzeuger aggregieren und tatsächlich vermarkten. In den hier betrachteten Fallbeispielen waren dies große Wirtschaftsunternehmen, aber auch Tochterunternehmen wissenschaftlicher Fachgesellschaften und andere wissenschaftsnahe Entitäten. Solche Betreiber können durch eine Forschungscommunity zur Vermarktung der Daten entweder „ermächtigt“ oder beauftragt sein oder aber den Dienst eigeninitiativ am Markt anbieten.

⁶³ Beispiele sind TAIR in den Lebenswissenschaften, PANGAEA in den Umweltwissenschaften sowie NOMAD in den Materialwissenschaften. Vgl. Darstellung im Anhang.

Praxisbeispiele 9

Im Fall des biologischen Datendienstes *TAIR* wurde die Betreibereinrichtung Phoenix Bioinformatics aus einem wissenschaftlichen Projektteam heraus gegründet, um den Betrieb des Dienstes nach Auslaufen der langjährigen Förderung aufrechtzuerhalten. Es handelt sich um ein Non-Profit-Unternehmen, das neben *TAIR* auch andere Services für die Community betreibt.

Auch die Dortmund Data Bank (DDBST) GmbH, eine der Betreiber von *Detherm*, der numerischen Datenbank für thermophysikalische Stoffdaten, ging 1989 aus einer wissenschaftlichen Arbeitsgruppe hervor. Der Dienst wird gemeinsam mit der Gesellschaft für Chemische Technik und Biotechnologie angeboten.

Das sozialwissenschaftliche *PsychArchives* wurde 1971 aus der Deutschen Gesellschaft für Psychologie heraus an der Universität Trier gegründet und später über das heutige Leibniz-Zentrum für Psychologische Information und Dokumentation verstetigt, dessen Mission speziell die Informationsbereitstellung ist.

Ein Beispiel für gezieltes Outsourcing der Betreiberrolle ist die Ausschreibung der Data Access and Information Services (DIAS) im europäischen Erdbeobachtungsprogramm *COPERNICUS*. Betreiberkonsortien können sich um Mittel zum Aufbau nutzerorientierter Dienste auf den vorhandenen Daten der europäischen Weltraumagentur ESA bewerben.

Für den rechtswissenschaftlichen Informationsdienst *Juris.de* suchte die öffentliche Hand als Datenerzeuger einen Verwertungspartner und gründete eine GmbH, deren Mehrheitseigner das heutige Bundesministerium für Justiz und Verbraucherschutz ist. Weitere Gesellschafter sind die Bundesrechtsanwaltskammer und – nach einer Teilprivatisierung – der Verlag Éditions Lefebvre Sarrut.

Anbieter eines Dienstes ist nicht gleich technischer Dienstleister

Mit dem Betrieb eines Dienstes sind unterschiedliche technische Aufgaben verbunden – vom Hosting einer Datenbank, Web-Schnittstellen bis hin zur Bereitstellung von Cloudspeicherplatz. Diese Aufgaben werden teilweise arbeitsteilig wahrgenommen, d.h. der Anbieter bzw. Betreiber eines Dienstes ist keineswegs immer mit dem **technischen Dienstleister** identisch.

Auch akademische Partner können die Aufgabe der technischen Betreuung übernehmen. Dies sind dann in der Regel deren Rechenzentren. Unter den Betreibern, die zugleich technische Dienstleistungen unter ihrem Dach vorhalten, sind zum Beispiel das FIZ Karlsruhe oder die Max Planck Digital Computing Facility. Einige Datenrepositorien im Wissenschaftsbereich werden auf einer kommerziellen technologischen Basis betrieben. In diesen Fällen ist der Zukauf von Speicherkapazitäten möglich.

Eine spezielle technische Dienstleistung ist die Sicherung der langfristigen Verfügbarkeit von Daten. Die Datenrepositorien *Dryad* und *Mendeley Data* kooperieren diesbezüglich mit dem niederländischen Data Archiving and Networked Services (DANS, einem Institut der Königlichen Niederländischen Akademie der Wissenschaften) – eine Kooperation, die dem gemeinsamen Angebot den Mehrwert hoher institutioneller Reputation garantiert. Auch wird das Verlags- und Bibliotheksnetzwerk CLOCKSS zur Langzeitarchivierung genutzt.⁶⁴

Kooperationen bei
Sicherung der Lang-
fristverfügbarkeit

Insgesamt kann auf Grundlage einer Analyse öffentlich zugänglicher Dokumente ein Outsourcing technischer Teilaufgaben bei zahlreichen Diensten lediglich vermutet werden: Angaben zur Frage der technischen Seite von Datendienstleistungen sind schwer bis gar nicht zu finden. Ein Konflikt über die Vergabe des wesentlichen Teils technischer Dienstleistung für 4TU.ResearchData in den Niederlanden an Figshare verdeutlicht den Anspruch von Non-Profit/Open Source Anbietern, einen Anteil des Marktes für sich zu gewinnen. Es zeigt aber auch deutlich die Limitationen, denen sie im Wettbewerb unterliegen.

Outsourcing
technischer
Teilaufgaben

Praxisbeispiele 10

Im Jahr 2020 hat das niederländische 4TU.ResearchData die technische Basis für Datenservices erneuert, die zuvor auf Basis eines gängigen Open-Source-Systems aus dem Publikationsbereich betrieben worden sind. Im Ausschreibungsverfahren erhielt der kommerzielle Anbieter *Figshare* den Zuschlag. Die Umstellung auf eine nun proprietäre Lösung erklärte das Konsortium mit Schwierigkeiten, die Open-Source-Software selbst weiterzuentwickeln. Die gesichteten Alternativen böten nicht die erforderliche Funktionalität bzw. könnten die Anforderungen der Ausschreibung nicht erfüllen. Die Non-Profit-Initiativen *Dryad* und *Zenodo* widersprachen dieser Darstellung öffentlich. Sie bemängelten Ausschreibungsbedingungen, durch welche sie praktisch aus dem Wettbewerb ausgeschlossen würden. Ihrer Einschätzung nach würden große kommerzielle Anbieter privilegiert. Dass Open-Source-Software unberechtigterweise als nicht nachhaltig gelte, würde die Überzeugung befördern, nur kommerzielle Lösungen großer Anbieter seien für die Wissenschaft geeignet.⁶⁵

Nicht selten finden sich Akteurskonstellationen, in denen einzelne Partner als eigenständige Vermittler entweder in der Akquise von Datennutzern („Vertrieb“) oder von Datenerzeugern auftreten. So ist beispielsweise das FIZ Karlsruhe europäischer Vertriebspartner für den US-amerikanischen Anbieter Chemical Abstract Services (CAS).

Broker zwischen
Datenerzeugern
und
Datennachfragern

⁶⁴ Neben Dryad auch durch Figshare: clockss.org/2012/04/figshare-preserves-with-the-clockss-archive/, zuletzt abgerufen am 2.06.2021.

⁶⁵ Vgl. zwei aufeinander bezogene Blogposts- Open Working (2020) – Why figshare? sowie Teal et al. (2020) – Alternatives Exist, zuletzt abgerufen am 2.06.2021.

Im Bereich der „Smart Cities“ werden Agenturen gegründet, die als Broker zwischen den datenerzeugenden Behörden einer Stadt und den Start-ups/App-Entwicklern als Datennachfrager vermitteln – im Sinne einer Hilfe, um Projekte aller Art und auch Forschungsprojekte zu entwickeln. Vergleichbare Agenturen zur Stimulation von Datennutzung sind in der Wissenschaft selten. In den Sozialwissenschaften hatte allerdings der RatSWD besonders in seinen Anfängen eine Art Broker-Rolle, indem er für die Wissenschaft mit Behörden über die Bereitstellung von für die Forschung attraktiven Daten verhandelt hat.

Praxisbeispiele 11

TriNetX, eine Plattform für klinische Studiendaten, agiert als Broker zwischen Einrichtungen, die Studien durchführen, und Pharmaunternehmen, also Auftraggebern. Technisch handelt es sich um eine Art „Datenmarktplatz“: Kunden laden Daten hoch, behalten aber ihre Verfügungsrechte und entscheiden, wem der Zugang gewährt wird.

Im Bereich der öffentlichen Hand will der *MobilitätsDatenMarktplatz* (betrieben von der Bundesanstalt für Straßenwesen) eine sichere Plattform für den Austausch von Mobilitätsdaten bereitstellen und baut um dieses Datenbankangebot herum ein Netzwerk auf.

Ähnlich wie Mendeley Data oder Researchgate sind soziale Netzwerke für die Kommunikation unter Wissenschaftlern, die teils um einen Dienst herum aufgebaut werden und vorher nicht in dieser Form existiert haben, Teil der Verwertungs- bzw. Vermarktungsstrategie der kommerziellen Anbieter.

3.2.3 Qualitätssicherung als Handlungsfeld: die Rolle des Kurators

Aufbereitung,
Pflege, Begutachtung
und Ergänzung als
eigenständiges
Handlungsfeld

Kuratierung umfasst die Aufbereitung, technische und inhaltliche Pflege der (Meta-)Daten sowie ihre Begutachtung und die Ergänzung des Datenbestands. Die Wahrnehmung dieser Aufgabe ist wegen ihrer besonderen Bedeutung für das wissenschaftliche Arbeiten mit qualitätsgesicherten Daten ein eigenständiges Handlungsfeld, für das – sofern in einem Dienst praktiziert – ein Akteur die Verantwortung übernehmen muss. Die methodengetriebene Metadaten-Kuratierung ist letztlich eine wesentliche Voraussetzung für die Erschließung und Findbarkeit der Daten, das heißt: für das Fundament qualitativ hochwertiger Wissenschaft.

Arbeitsteilige
Kuratierung zwischen
Datenerzeuger und
Betreiber

Oft wird die Kuratierung arbeitsteilig zwischen Datenerzeuger und Betreiber organisiert (von sehr akribisch bis eher cursorisch). Teils wird die Aufgabe aber auch an die Datenerzeuger oder ihre Fachgemeinschaft/Community zurückgegeben. So bietet *ChemSpider* ähnlich einem Wiki angemeldeten Nutzern die Möglichkeit, die Korrektheit hinterlegter Informationen zu überprüfen und ggf. zu ergänzen. Die meisten der generischen Long-Tail-Repositoryen überlassen die Kuratierung weitgehend den Datenerzeugern.

Soweit es sich um institutionelle Datenrepositorien handelt, nehmen ggf. Bibliotheken Aufgaben der inhaltlichen Erschließung wahr. Andere Betreiber bieten das Kuratieren (aggregieren, annotieren etc.) als wesentlichen Teil ihrer anbieterseitigen Wertschöpfung an, sie agieren wie ein klassischer Verlag im Publikationswesen (teils gegen Entgelt). Kuratierung ist auch ein essentielles Merkmal von Referenzkorpora. In der Chemie sind die Fachgesellschaften diesbezüglich wichtige Akteure; im Bereich der Literaturwissenschaften sind dies einerseits Einrichtungen wie Akademien der Wissenschaften, andererseits Fachbereiche an Universitäten, die digitale editionsphilologische Schwerpunkte entwickeln. Hierfür wird – wie bei den Betreibern der Fallbeispiele aus der Chemie – auch eigenes Personal vorgehalten.

3.3 GOVERNANCE – ZWISCHEN MARKTPLATZ- UND REDAKTIONS MODELL

Datendienste für die Wissenschaft sollen ein für Forschung, Lehre und Transfer leicht nutzbares Leistungsspektrum bieten. Um dieses zu gewährleisten und die Zugänge hierzu zu sichern, ist es wichtig, dass die wissenschaftlichen Beteiligten angemessen in die Ausgestaltung und Weiterentwicklung der Dienste involviert sind. Hierbei können die Akteure in ganz unterschiedlichen Rollen auftreten, die in zahlreichen Fällen auch konvergieren: am augenscheinlichsten in der Doppelrolle als Datenerzeuger und Datennutzer, zusätzlich als Anbieter, Betreiber, Datensubjekt, Kurator etc.

Ausgestaltung und Weiterentwicklung der Dienste

Unter den vom RfII betrachteten Fallbeispielen lassen sich zu dem Aspekt der Einbindung bzw. Mitwirkung wissenschaftlicher Akteure unterschiedliche Ausprägungen finden. Von kundenorientierten Monopolisten bis zu basisdemokratischen Graswurzel-Initiativen ist das Spektrum der praktizierten Governance-Regimes breit gefächert. Mit Blick auf die genuin aus der Wissenschaft heraus (und für die Wissenschaft) entwickelten Dienste bietet es sich für einen Vergleich an, von einem Konzept der Mandatierung für unterschiedliche Formen der Governance und des Betriebs eines Dienstes zu sprechen. Subjekt der Mandatierung ist häufig eine Fachgemeinschaft – in der Regel repräsentiert durch eine Fachgesellschaft – oder durch eine für Infrastrukturbedarfe gegründete wissenschaftliche Einrichtung. Auch die Wissenschaftspolitik und die Forschungsförderung setzen Anreize, die zur Formierung von Akteuren oder Akteurskonstellationen führen, die zum Zweck der Programmausführung mandatierte Einrichtungen und Dienste gründen. Letztlich kommen hier auch Einrichtungen für die generische Infrastrukturentwicklung wie Universitätsbibliotheken und Rechenzentren ins Spiel. Für die Frage der Mandatierung spielt es keine Rolle, ob ein Dienst sich aus öffentlichen Mitteln speist oder kommerziell betrieben wird. Wissenschaftliche Akteure können, wenn sie hierzu mandatiert sind, problemlos kommerzielle Anbieter als Intermediäre

Breitgefächerte Governance-Regimes

einsetzen, wenn die Funktionalität eines Dienstes für die Wissenschaft so am besten gewährleistet werden kann.⁶⁶

Mandatierung über internationale Vereinbarungen

Bei einigen Datendiensten sind Auswahl und Kuratierung der Inhalte auch über internationale Vereinbarungen mandatiert. Akteure handeln in transnationalen Netzwerken oder europäischen Infrastrukturen Standards und Anforderungen für übergreifende Dateninfrastrukturen aus. Grundlage für die Übertragung der Daten sind hier teilweise internationale Übereinkommen, die auf ministerialer Ebene abgeschlossen werden. Im Fall der europäischen Forschungsinfrastrukturen der ESFRI-Roadmap geht dem ein wissenschaftlicher Bottom-up-Prozess voraus, in dem aus Fachgemeinschaften und Forschungs-Communities heraus Infrastrukturbedarfe artikuliert und für die Umsetzung bereits bestehende oder neu zu gründende Einrichtungen mandatiert werden.

Stark abstrahierend ließen sich aus den vom RfII betrachteten Fallbeispielen zwei gänzlich unterschiedliche Governance-Regimes modellieren, die hier im Folgenden idealtypisch skizziert werden.

3.3.1 Das Forum- oder Marktplatzmodell

Auslagerung der Qualitätssicherung und Orientierung an Nutzungszahlen

Wo Dienste eher auf technische Leistungen hin orientiert sind – eine Plattform für die Publikation anbieten oder Informationsressourcen aggregieren –, sind die Entscheidungen über Art und Umfang der Inhalte nachfrageorientiert. Die redaktionelle Leistung bzw. inhaltliche Qualitätssicherung wird oftmals den Nutzerinnen und Nutzern selbst überlassen bzw. (vom Dienst her gesehen) an Experten ‚nach außen‘ gegeben. Entweder sind die Datenerzeuger selbst verantwortlich, oder externe Kuratoren (der zuliefernden Ressourcen) kümmern sich um Inhalte und stabile Prozesse. Der Betrieb wird über die Nutzungsbedingungen, im Einzelfall auch über Nutzungsverträge geregelt. Ziel des Betriebs sind hohe Nutzungszahlen und ein Maximum an „traffic“.

Ausrichtung an den Bedarfen (zahlender) Datennachfrager

Die Ausrichtung an den Bedarfen (zahlender) Datennachfrager ist für einige der kommerziellen Datendienste handlungsleitend. So erfolgt z.B. Im Fall der

⁶⁶ In seinem Positionspapier zur Datenqualität betont der RfII allerdings das besondere Problem der Nutzung proprietärer Hard- und Software im wissenschaftlichen Forschungsprozess, die den Forscherinnen und Forschern die Kontrolle über die Prozessierung von Daten entzieht und zur Replizierbarkeitskrise beiträgt. Der RfII unterstreicht die Bedeutung wissenschaftseigener Open-Source-Software bzw. der Transparenz des Quellcodes. Der RfII empfiehlt den Wissenschaftsorganisationen, sich zusammenzuschließen. Wo möglich sollten wissenschaftseigene Lösungen entwickelt werden und – wo nötig – gegenüber den kommerziellen Herstellern eine Nachvollziehbarkeit der internen Datenverarbeitungsprozesse, des Codes und der Algorithmen eingefordert werden. RfII (2019) – Herausforderung Datenqualität, Empfehlung 4.3.4, S. 87.

Discovery-Plattform *Ontosight* eine umfassende Aggregation vorhandener offener Daten aus Patent-, Literatur- und Sequenzdatenbanken. Um das Vertrauen der Nutzer wird mit (personeller und technischer) Expertise sowie in einigen Fällen mit der Herkunft der Daten (öffentliche Hand und Wissenschaft) geworben.

Im Fall der Repositorien für Long-Tail-Daten findet sich zwar eine gewisse Orientierung an Nutzerbedarfen, allerdings stehen hier die Datenerzeuger und ihr Ablage- und Archivierungsinteresse im Mittelpunkt. In der Konsequenz erfolgt oft keine oder nur eine rudimentäre technische Prüfung der Einreichungen vor Veröffentlichung, da diese Repositorien eher auf maximale Akquise ausgelegt sind.

3.3.2 *Das Redaktions- oder Editionsmodell*

Andere Dienste haben den Charakter wissenschaftlicher Sammlungen nach transparenten Ordnungs- und Qualitätsgesichtspunkten oder sie bilden sich als Infrastruktur um bedeutsame wissenschaftliche Erhebungen herum (z.B. Panelstudien). Sie sind durch ein nachvollziehbares wissenschaftliches/fachliches Interesse begründet und entsprechend eng mit dem Forschungshandeln verknüpft. Als Instrumente sind hier zum Beispiel fachlich geprägte Kuratierungskonzepte oder Studienpläne wichtig. Auch sind in der Governance-Struktur qualitätssichernde Organe wie wissenschaftliche Beiräte vorgesehen. Während diese Form der Governance darauf abzielt eine gute Datenqualität sicherzustellen, kann die Bindung an konkrete Forschungsfragestellungen auch dazu führen, dass eine Engführung von Inhalten oder Akteuren erfolgt. Entsprechend limitiert sind Zielgruppe und „traffic“. Auch können die Kosten für den Betrieb eines solchen Dienstes sehr hoch sein, soweit die Kuratierung nicht kollaborativ (sowie durch ehrenamtlichen Einsatz von Schlüsselpersonen) erfolgen kann. Deshalb werden solche Dienste zumindest in Deutschland dort, wo sie sich als strukturbildend für ganze Fachgemeinschaften erweisen, nach eingehender Evaluierung institutionell verselbständigt oder in institutioneller Trägerschaft einer Forschungseinrichtung – häufig eines Leibniz-Instituts oder im Bereich der historischen Geistes- und Kulturwissenschaften im Rahmen des nationalen Akademienprogramms – verstetigt bzw. an den Akademien mit Projektlaufzeiten versehen, die einer Verstetigung nahekommen.

Für diese Einrichtungen wird bzw. ist die Verwendung des eigenen institutionellen oder institutionell betreuten Datenbestands handlungsleitend, d.h. es werden in der Regel keine Inhalte externer Datenerzeuger für Zwecke der Nutzung und Verwertung aufgenommen. Dies gilt auch für viele Dienste öffentlich finanzierter Gedächtniseinrichtungen oder die Datenintegrationszentren der Medizininformatik-Initiative.

Fachlich geprägte
Kuratierungskonzepte
i.S. von Sammlungen/
Erhebungen als
Kristallisationspunkte

Kuratierung
institutioneller
bzw. institutionell
betreuter
Datenbestände

Nachhaltigkeits-
probleme
projektfinanzierter
Dienste

Redaktionell bzw. editorisch gelenkte Dienste werden üblicherweise dort nachhaltig betrieben und weiterentwickelt, wo öffentliche Zuwendungsgeber zur Finanzierung beitragen und technische Voraussetzungen existierender Trägereinrichtungen ‚mitgenutzt‘ werden. Dienste dieses Typs haben häufig ein Nachhaltigkeitsproblem, wenn sie nicht institutionell verstetigt, sondern projektfinanziert sind.⁶⁷ Aus wissenschaftspolitischer wie auch aus wirtschaftlicher Sicht bauen sich im Rahmen solcher Dienste gleichwohl teilweise hohe „Werte“ auf, die nicht nur für Forschungsgemeinschaften/Communities wichtig sind, sondern teils auch auf dem ‚freien Markt‘ auf Interesse stoßen (würden).

Reputationswirksame
Begutachtungs- und
Auswahlverfahren

Im Bereich der fachlichen und wissenschaftsnahen Repositorien in den Sozial-, Verhaltens- und Wirtschaftswissenschaften praktiziert das sozioökonomische Datenzentrum *SEDAC* ein wissenschaftsgeleitetes Begutachtungs- und Auswahlverfahren, das dem der wissenschaftlichen Zeitschriften ähnelt. Die Aufnahme eines Datensatzes in den Dienst ist entsprechend mit wissenschaftlicher Reputation verbunden. Gleichwohl ist dieser Ansatz ein Sonderfall unter den Datenrepositorien. Bei *Dryad* oder *SoWiDataNet*|*Datorium* gibt es zwar Personal, das technische Aspekte, Lizenzen und Personenrechte betreffende Prüfungen der Daten bzw. eine inhaltliche Prüfung der Metadaten durchführt. Es existiert jedoch kein mit *SEDAC* vergleichbares Sammlungskonzept. Andere Repositorien beschränken die Aufnahme auf Daten aus registrierten Studien (*PsychArchives*) oder bereits begutachteten Daten (*Mendeley Data*).

Praxisbeispiele 12

Naheliegende Beispiele für das Redaktions-/Editionsmodell sind das *Digitale Wörterbuch der deutschen Sprache* oder das *Deutsche Textarchiv*. Beide Dienste sind Langzeitprojekte an der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW). Die redaktionelle Zuständigkeit für die Korpora liegt bei jeweils mehrköpfigen sogenannten Projektgruppen der BBAW, die über Ergänzungen weiterer zu aggregierender Textkorpora bzw. Texte nach festgelegten wissenschaftlichen Kriterien entscheiden.

Wissenschaftliche Qualitätssicherung erfolgt auch bei den Referenzdatenbanken in der Chemie (*CAS*, *DETERM*) und der Geologie (*World Stress Map*). Die verwerteten bzw. vermarkteten Inhalte werden von qualifizierten Experten erstellt (quasi „herausgegeben“) und haben einen hohen Anspruch an geprüfte Qualität.

Dienste wie *CAS* in der Chemie oder auch *TAIR* in der molekularen Pflanzenforschung beschäftigen fachlich ausgebildete Redakteure, die für den Dienst geeignete Inhalte auswählen und den Kontakt zur Domäne halten. Die Datennachfrager sind zahlende Kunden, die Verträge mit dem Betreiber schließen. Diese Dienste arbeiten unter hohen Qualitätsstandards und im Austausch mit wissenschaftlichen Fachgemeinschaften.

⁶⁷ Vgl. Empfehlung 1 in RfII (2016) – Leistung aus Vielfalt.

Die aufwendige Redaktion verursacht hohe Kosten, die sich in entsprechenden Gebühren widerspiegeln. Als Non-Profit-Einrichtungen erzielen diese Dienste mit den Gebühren allerdings keinen Gewinn – Gebühren decken hier lediglich die Kosten für Betrieb und Weiterentwicklung.

3.3.3 Einbindung der Nutzer in die Governance

Die Einbindung der wissenschaftlichen Nutzerinnen und Nutzer von Datendiensten, die als eigenständige wissenschaftliche Einheiten oder im Rahmen von Forschungsinstituten verstetigt sind, gilt aus wissenschaftlicher Sicht als obligatorisch und ist ein wichtiges Element der Qualitätssicherung. Entsprechend muss auch das Vorhandensein aktiver Nutzerbeiräte in der Governance von Datendiensten gesichert sein und im Rahmen periodischer Evaluationen durch externe wissenschaftliche Gutachtergruppen überprüft werden können. Ein häufig nicht reguliertes und damit auch unbekanntes Feld stellen in dieser Hinsicht bislang die meisten Datenrepositorien an Universitäten und Universitätsbibliotheken dar.

Qualitätssicherung
durch externe
Evaluation und
Nutzerbeiräte

Bei den von der öffentlichen Hand betriebenen Datendiensten sind die Datenauswahl und Kuratierung oft nutzerorientiert gestaltet und Nutzer sind auch als Mitgestalter beteiligt. So sind die behördlichen Forschungsdatenzentren in Aushandlungsprozessen mit der Wissenschaft entstanden. Der RatSWD (paritätisch besetzt aus Wissenschaft und Behörden) fungierte hier als eine Art Beirat, der die Erschließung der fraglichen Daten über das Modell Forschungsdatenzentren aktiv betrieb.⁶⁸ Eine flankierende Akkreditierung stützt diese Dienste mit einem Renommee aus, das mit Reputation im Wissenschaftssystem verglichen werden kann.

FDZ als Beispiele

Ein Element mit vergleichbarer Funktion können Nutzerbeiräte in den Datendiensten sein. So verfügt das bayerische Portal *bavarikon* über einen Beirat der beitragenden Datenerzeuger (hier: Gedächtniseinrichtungen). Das Datenrepositorium *Dryad* praktiziert ein Mitgliedschaftsmodell, hier bestimmen Forschungsorganisationen und Verlage über das Board of Directors und die Satzung.

Nutzerbeiräte und
Mitgliedschaften

Auch bei marktorientierten Diensten können Beiräte sowohl wissenschaftliche wie auch nichtwissenschaftliche Mitglieder haben. Im Fall von *TAIR* bzw. dessen non-profit orientierter Betreibergesellschaft Phoenix Bioinformatics bestimmt ein Aufsichtsgremium („Board“) die Geschäftsstrategie mit, das mit Vertreterinnen und Vertretern der Wissenschaft und der Datenwirtschaft besetzt ist.

⁶⁸ KVI (2001) – Wege zu einer besseren informationellen Infrastruktur (Langfassung), 37 f.

3.4 ZUGANGSREGIMES

Zugang zu bestimmten Leistungen oder Leistungsqualitäten als Grundlagen der Geschäftsmodelle

Der Zugang zu Forschungsdaten umfasst das Finden und den gelingenden Zugriff. Zugang setzt Bereitstellung plus Zugänglichmachen (z.B. eine maschinenlesbare oder von Personen bedienbare Schnittstelle) voraus – das „Accessible“ im Sinne des FAIR-Prinzips). Die Europäische Charta für den Zugang zu Forschungsinfrastrukturen definiert „Zugang“ dagegen breiter, nämlich als „den legitimen und autorisierten physischen, entfernten und virtuellen Zugang zu Forschungsinfrastrukturen, die Interaktion mit ihnen und ihre Nutzung sowie auf die von Forschungsinfrastrukturen den Nutzern angebotenen Dienste.“ Ein solcher Zugang könne beispielsweise gewährt werden „zu Maschinenzeit, Rechenressourcen, Software, Daten, Datenkommunikationsdiensten, Vertrauens- und Authentifizierungsdiensten, Probenvorbereitung, Archiven, Sammlungen, Aufbau, Durchführung und Abbau von Experimenten, Aus- und Weiterbildung, Expertenunterstützung und Analysediensten.“⁶⁹ Viele, aber nicht alle der hier aufgeführten Ausprägungen von „Zugang“ finden sich heutzutage in den von der Wissenschaft genutzten Datendiensten. Zugangsbarrieren haben oft wirtschaftliche Gründe. Denn dass die differenzierte Handhabung des Zugangs zu bestimmten Leistungen oder Leistungsqualitäten zu den Grundlagen der Geschäftsmodelle der Angebote gehört, liegt auf der Hand.

Zugänglich, aber nicht „offen“

In fast allen Domänen sind Datenressourcen zwar öffentlich verfügbar, aber nicht durchweg „offen“ im Sinne der netzpolitischen Open Definition⁷⁰. Die meisten Dienste erfordern mindestens eine Registrierung, zudem gelten zahlreiche rechtliche und technische Einschränkungen. So können Download-Möglichkeiten gesperrt sein oder der Dienst verfügt über keine Schnittstelle zum Harvesting der Metadaten. Diese Einschränkungen sind teils begründet, teils nicht immer leicht nachvollziehbar. Hinzukommen Bezahlschranken auch für „offene“ Inhalte, zum Beispiel die Subskription als Voraussetzung für die Nutzung eines Dienstes oder die Erhebung von Gebühren bzw. Entgelten. Auch der Zugang zu Metadaten ist keineswegs immer vollumfänglich gegeben. Gedächtniseinrichtungen machen in manchen Fällen auch für die von ihnen vergebenen Metadaten Urheberrechte geltend.

⁶⁹ opendefinition.org/od/2.1/de, zuletzt abgerufen am 2.06.2021.

⁷⁰ „Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness).“ (The Open Definition, 2015, opendefinition.org, zuletzt abgerufen am 2.06.2021).

3.4.1 Zugriff nach Registrierung und auf Antrag

Für den „gelingenden Zugriff“ besteht in vielen Diensten eine **Registrierungspflicht**, d.h. die Datennachfrager müssen sich dem Betreiber gegenüber identifizieren und geben die dazu geforderten Daten in eine Datenbank des Anbieters und des Betreibers ein. Teils hat dieses Identifikationserfordernis mit der Kostenpflichtigkeit der Dienstleistungen zu tun, teils mit der Art der Daten, z.B. weil ein Zugang zu Inhalten gewährt wird, deren Nutzung rechtlichen Einschränkungen unterliegt. Hierzu gehören der Datenschutz, z.B. die Europäische Datenschutzgrundverordnung und der Schutz geistigen Eigentums (Urheberrecht, Markenrecht, Patentrecht).

Rechtliche und monetäre Gründe für Identifikationserfordernis

Für urheberrechtlich geschützte Objekte gilt ein Forschungsprivileg, das im Urheberrecht verankert ist. Bei Zugriff auf diese Ressourcen ist es daher erforderlich, sich als Angehöriger des Wissenschaftssystems auszuweisen. Ein Beispiel dafür ist das *Deutsche Referenzkorpus (DeReKo)*. Registrierten Datennachfragern kann hier mithilfe von über 200 Lizenzvereinbarungen die Nutzung der Texte für nicht-kommerzielle bzw. Forschungszwecke ermöglicht werden.

Forschungsprivileg bei urheberrechtlich geschützten Objekten

In den gesichteten Diensten werden für die technische Organisation des Zugangs per Registrierung oftmals existierende, transnationale Infrastrukturen des Wissenschaftssystems genutzt, wie zum Beispiel eduGAIN. Dabei identifizieren und authentifizieren sich die Nutzer über eine Forschungseinrichtung. Andere Anbieter führen eigene Nutzer-Accounts, die ggf. an Kundinnen und Kunden (auch kommerzielle) oder nur an bestimmte Nutzergruppen vergeben werden (Angehörige bestimmter Sektoren oder Institutionen).

In Domänen wie der Medizin, den Sozial- und Verhaltenswissenschaften, der Zeitgeschichte oder der Ethnographie sind archivierte Daten oftmals personenbezogen und somit besonders schützenswert. Zugang kann in diesen Fällen erst auf **Antrag** gewährt werden, wobei zum Beispiel ein legitimes wissenschaftliches Interesse nachgewiesen werden muss. Gewerbliche Nutzung kann zudem explizit ausgeschlossen sein. Der Zugang ist also adressaten- und zweckgebunden.

Legitimes wissenschaftliches Interesse als Grundlage für Zugang auf Antrag

Der Rat für Sozial- und Wirtschaftsdaten (RatSWD) hat sein Akkreditierungsmodell auf die Etablierung eines standardisierten Zugangsmodells ausgerichtet, das der Forschung transparente und verlässliche Bedingungen bieten soll. Datennutzungsverträge sind ein übliches Instrument, um die Forschenden zur Geheimhaltung zu verpflichten. Der praktische Zugang erfolgt über sog. „safe rooms“, in denen primäre Daten (Mikrodaten) nur vor Ort oder über eine Remote-Verbindung eines anderen Standortes angesehen bzw. ausgewertet werden können. Für die ortsungebundene Bereitstellung werden gesonderte Datenprodukte von unterschiedlichen Anonymisierungsgraden erstellt (sog. Scientific Use Files und Campus Files mit einer gesteigerten faktischen Anonymisierung sowie die absolut anonymisierten sog. Public Use Files).

Transparente und verlässliche Bedingungen durch Standardisierung vertraglicher Regelungen

Für die künftigen Datenintegrationszentren der Universitätsklinik, die im Rahmen der Medizininformatik-Initiative entstehen und ebenfalls hochsensible Daten vorhalten, ist ebenfalls ein Antragsverfahren vorgesehen. Über die Anträge entscheiden jeweils lokale „Use and Access Committees“, so dass ein Antrag, der sich auf Daten mehrerer Zentren bezieht, ggf. mehrere Genehmigungsschritte durchlaufen muss.⁷¹ Zudem wird eine Koordinations- und Registerstelle (ZARS) geschaffen, die ein einheitliches Antragsverfahren und die Durchführung der Datennutzung zentral verwaltet und gegenüber Bürgern und der Fachöffentlichkeit Transparenz über die Zwecke und Erfolge der gewährten Datennutzung schaffen soll.

3.4.2 Offener Zugang zu Inhalten

Offener Zugang („anyone can freely access“) ist nur dort möglich, wo keinerlei rechtliche Einschränkungen für die Datennutzung, -verwertung und -vermarktung bestehen, wo Daten also frei lizenziert angeboten werden.

Bereitstellung
offener Daten
qua ehrenamtlicher
Kuratierung

Er kommt auch in Bereichen vor, in denen Daten einen hohen Marktwert haben. Im Fall von *ChemSpider* liegt beispielsweise der Betrieb in der Verantwortung einer Fachgesellschaft, ist durch Engagement der Fachgemeinschaft und vielseitige Finanzierung flankiert. Dies ermöglicht es, kuratierte Inhalte kostenfrei zur Verfügung zu stellen. Eine Studie aus den Lebenswissenschaften zeigt auf, wie die öffentlich finanzierten Dateninfrastrukturen in diesem Bereich die Entstehung kleiner und mittlerer Unternehmen stimuliert haben, die Wertschöpfung in Form von bioinformatischen Dienstleistungen betreiben. Die Autoren verweisen zudem auf Studien zu Geodaten und Open Government Data, die ähnliche Effekte aufzeigen.⁷²

Sektorübergreifender
Zugang zu
hochwertigen Daten
auf Grundlage
politischer
Entscheidungen

Im Fall der Erdbeobachtungsdaten sind die europäischen Mitgliedstaaten als Geldgeber an dem Wissenstransfer und der dadurch ermöglichten Innovation interessiert und fördern selbst die Verwertung der Daten durch offenen Zugang. Die Europäische Kommission hat Interesse daran, auch weitere sog. High Value Data uneingeschränkt sektorübergreifend zur Entwicklung und Vermarktung neuer Dienste zugänglich zu machen und will erheblich in sog. Datenpools investieren.⁷³

Teiloffenheit im
Freemium-Modell

Bei anderen Diensten ist offener Zugang teils auf Metadaten beschränkt bzw. auf einen Teil des Angebots, wie bei den Forschungsdatenzentren mit ihren Public Use Files. Bei Bilddaten kann der offene Zugang nur eine gering aufgelöste (und

⁷¹ medizininformatik-initiative.de/de/nutzungsordnung, zuletzt abgerufen am 2.06.2021.

⁷² ELIXIR (2017) – Public Data Resources.

⁷³ EC (2020) – Eine europäische Datenstrategie.

damit schlecht nutzbare) digitale Version des Werks betreffen oder Derivate der digitalen Objekte (Auszüge oder Kleinformate, wie Inhaltsverzeichnisse, Vorschaubilder, Audio- und Videoausschnitte). Solche „teiloffenen“ Zugänge finden sich insbesondere in Diensten, die ein Freemium-Modell praktizieren sowie dort, wo ein Betreiber Daten eines unternehmerisch agierenden Dritten als Teil von dessen Verwertungskette aggregiert (zum Beispiel übergreifende Bildarchive bzw. Kataloge, die Findinstrumente auch für oder überwiegend für zahlungspflichtige Angebote bereitstellen). Die Rechte an den Datendiensten liegen dann dauerhaft beim Datenerzeuger, der sie unter mehr oder weniger offenen Nutzungsbedingungen verwertet (Wissenschaft) oder vermarktet (Wirtschaft).

3.5 FINANZIERUNG VON DIENSTEN, MONETARISIERUNG VON DATEN SOWIE DATENBEZOGENEN LEISTUNGEN

Die fünfte Vergleichsdimension für Datendienste ist das jeweilige Geschäftsmodell. Insbesondere die Verlässlichkeit der Einnahmegenerierung ist entscheidend dafür, ob ein Dienst langfristig planen und sein Portfolio ggf. ausbauen oder Leistungen verbessern kann. Im Punkt der (Re-)Finanzierung treten Zieldifferenzen zwischen wirtschaftlichen und wissenschaftlichen Motiven zum Betrieb eines Dienstes besonders deutlich zutage: Oft sind wirtschaftlich kurzfristig durchaus attraktive Geschäftsmodelle aus wissenschaftlicher Sicht nicht nachhaltig; nicht selten sind aber auch wissenschaftlich plausible, auf der Basis allmählicher Refinanzierung (oder dank Fördergeld) attraktive Geschäftsmodelle wirtschaftlich gesehen auf Dauer nicht rentabel (genug). Die in der Fallbeispielanalyse des Rfll gefundenen Ausprägungen von Finanzierungen der Dienste zeigt die folgende Tabelle.

Verlässlichkeit der Einnahmegenerierung als Grundlage der Geschäftsmodelle

Tabelle 3: Einnahmequellen und ihr Einfluss auf die Perspektive eines Dienstes

Ausprägungen von Finanzierung
Öffentlich/institutionalisiert („im Rahmen des Auftrags“) = stetige Bestandssicherung/ Weiterentwicklung
Öffentlich/projektgebunden (Drittmittel) = prekärer Bestand/Weiterentwicklung bei Anschlussprojekt
Mitgliedsbeiträge = Bestand und Weiterentwicklung abhängig von Größe und Commitment der Community
Spenden = Bestand und Weiterentwicklung abhängig von Größe und Commitment der Community und weiterer Unterstützer
Subskriptionsgebühren für Informationsangebote = Bestand und Weiterentwicklung nachfragegesteuert
Einnahmen aus Verkauf oder Lizenzierung von Einzelobjekten bzw. Technologien (etwa „white label use“ oder Zubuchen von Applikationen) = Bestand und Weiterentwicklung nachfragegesteuert
Einnahmen aus Deckungsbeiträgen für die Kosten der Veröffentlichung/Speicherung (contributor-pays Modell) = Bestand und Weiterentwicklung nachfragegesteuert
Einnahmen aus Dienstleistungen wie Beratung oder Kuratierung = Bestand und Weiterentwicklung nachfragegesteuert

Die vom RfII betrachteten Fallbeispiele zeigen eine Vielfalt an Geschäftsmodellen, die sich zunächst einmal nicht wesentlich vom Publikationssektor unterscheiden: Subskription, Varianten von „contributor pays“ oder „Freemium“-Angebote (= Mix aus kostenfreien Basisangeboten und zusätzlich buchbaren kostenpflichtigen „Premiumangeboten“).

Monetarisierung von Leistungen und Leistungsportfolios

Konkurrierende und auf Monetarisierung von Leistungen angelegte Angebote finden sich primär im Bereich der generischen, also fachübergreifenden Datenrepositorien. Hier ist auffällig, dass sich für ein recht ähnliches Leistungsportfolio ganz unterschiedliche Erlösmodelle finden. Der Einsatz von Venture Capital und Crowdfunding kommt dagegen bislang sehr selten zum Tragen.

Im Folgenden werden fünf Typen der Finanzierung unterschieden:

- Zuwendungen aus institutioneller Finanzierung und Projektmitteln,
- Community-Support über Mitgliedsbeiträge und Spenden,
- Monetarisierung von Leistungen,
- Freemium-Modelle,
- Investments: Venture Capital und Crowdfunding.

3.5.1 Zuwendungen aus öffentlicher Finanzierung und Community-Support

Dort, wo Daten ohne Gebühren und Entgelte bereitgestellt werden, sollen Inhalte der Allgemeinheit zur Verfügung stehen. Die Kosten für den Betrieb solcher Dienste werden in der Regel von **öffentlichen Zuwendungsgebern** übernommen. Zu dieser Gruppe gehören zum einen die institutionellen Forschungsdatenrepositorien und Forschungsdatenzentren, aber auch transnationale Dateninfrastrukturen, wie z.B. die *Worldwide Protein Data Bank (wwPDB)*. Die Services der Collaborative Data Infrastructure *EUDAT* werden im Rahmen einer Partnerschaft von den teilnehmenden europäischen Rechenzentren kostenfrei angeboten.

Aufbau von Diensten durch (wiederkehrende) Projektfinanzierungen

Wiederkehrende **Projektfinanzierungen** zum Auf- oder Ausbau von Features sind Teil des Geschäftsmodells. Dies können in Deutschland z.B. bundes- oder landesweite Programme zur Digitalisierung der öffentlichen Verwaltung sein, Forschungsförderung oder auch EU-Programme, wie im Fall des *COPERNICUS* Erdbeobachtungsprogramms.

Das *COPERNICUS* Programm wird durch die europäischen Mitgliedstaaten finanziert, was auch die Verfügbarmachung der Daten durch die European Space Agency ESA beinhaltet – explizit auch mit dem Ziel der gewerblichen Nutzung bzw. Vermarktung. Es handelt sich um Big Data, die beim Nutzer eine bestimmte IT-Infrastruktur bzw. Rechenkapazitäten erfordern. Die Vermarktung der Daten

umfasst daher oftmals auch technologische Angebote wie Software-as-a-Service und Infrastructure-as-a-Service. Der Aufbau dieser Vermarktungsplattformen (oder „Marketplaces“) für Daten und Dienste wird mit europäischen Fördermitteln unterstützt.⁷⁴

Die Fallbeispiele *Worldwide Protein Bank (wwPDB)* sowie die Forschungsdatenzentren der sozialwissenschaftlichen Langzeitstudien *SOEP* und *NEPS* illustrieren exemplarisch, wie Dienste nach langjähriger Projektfinanzierung durch einen zentralen Förderer in eine dauerhafte öffentliche Förderung überführt werden können.⁷⁵ Die Überführung in staatliche Dauerfinanzierung ist jedoch nicht der einzige Weg zur Verstetigung. Die Fallbeispiele *Dryad*, *prometheus* und *TAIR* stehen für eine Verstetigung durch Community-Beiträge und Monetarisierung von Daten.

Erfolgreiche
Verstetigungen

Praxisbeispiele 13

Die *Worldwide Protein Bank (wwPDB)* wurde seit der Gründung in den 1970er Jahren überwiegend durch die US-amerikanische National Science Foundation finanziert. Nach dem endgültigen Auslaufen der Förderung übernahm 2003 ein internationales Konsortium aus vier Regionaldatenbanken den Betrieb. Diese erhalten jeweils finanzielle Unterstützung durch für sie zuständige Forschungsförderer. Der Betrieb wurde also in eine kollektive internationale Verantwortung überführt, was dem Charakter der Nutzergemeinschaft Rechnung trägt. Alle Daten stehen unter eine CC-0 Lizenz kostenfrei zur Verfügung und werden nicht nur für Erkenntnisgewinn in Forschung und Entwicklung, sondern auch für spezialisierte Informationsdienstleistungen Dritter nachgenutzt.

Das *Nationale Bildungspanel (NEPS)* und das *Sozio-oekonomische Panel (SOEP)* sind sozialwissenschaftliche Langzeitstudien mit umfangreichen und kontinuierlichen Erhebungen, die über eigene Datenzentren verfügen. Beide wurden nach dem Auslaufen der Projektförderung aufgrund ihrer überregionalen Bedeutung und des gesamtstaatlichen Interesses in die gemeinsame Wissenschaftsförderung von Bund und Ländern überführt. Die zugehörigen Datenzentren werden von Instituten der Leibniz-Gemeinschaft betrieben, die Daten auf Antrag für die wissenschaftliche Forschung durch die Wissenschaft national und international kostenfrei bereitstellen (gewerbliche Nutzung ausgeschlossen).

⁷⁴ Zuletzt 2018 mit der Ausschreibung von fünf so genannten DIAS, or Data and Information Access Services, copernicus.eu/en/access-data/dias, zuletzt abgerufen am 2.06.2021.

⁷⁵ In der Übernahme einer dauerhaften Verantwortung für den Betrieb eines Datendienstes durch Forschungseinrichtungen oder von ihnen gebildete Verbünde sieht der Wissenschaftsrat die Möglichkeit einer strategischen Positionierung und Profilbildung sowie die Möglichkeit zur Übernahme von Führerschaft bei Prozessen nationaler und internationaler Standardbildung. WR (2020) – Zum Wandel in den Wissenschaften, S. 49. Die Fallbeispiele bestätigen, dass die Verantwortungsübernahmen für einen für eine definierte Fachgemeinschaft grundlegenden Dienst Reputation und weitere Forschungsvorhaben mit sich bringt.

Einige Dienste akquirieren **Mitgliedsbeiträge** und **Spenden** von **wissenschaftlichen Communities**, die teilweise in Kombination das Geschäftsmodell bilden. Die Erhebung von Mitgliedsbeiträgen erfordert eine starke Bindung der beteiligten Stakeholder und von deren „Mitstreitern“, die das gemeinsame Ziel durch regelmäßige Beiträge unterstützen. Mit der Mitgliedschaft sind dann vereinfachte Nutzungsbedingungen und Teilhabe an Mitbestimmungsprozessen verbunden.

Praxisbeispiele 14

Das generische Repositorium *Dryad* wurde in der Aufbauphase durch die National Science Foundation (2007–2019) und verschiedene europäische Mittel gefördert, zunächst als ein Repositorium für Daten aus den Umwelt- und Lebenswissenschaften. Mittlerweile wurde es in einen gemeinnützigen Verein nach US-amerikanischem Recht umgewandelt und für Daten aus dem gesamten Bereich „Science“ geöffnet. Mitglieder sind wissenschaftliche Einrichtungen und Fachgesellschaften weltweit. Zum Finanzierungsmodell gehört die Partnerschaft mit Fachzeitschriften bzw. Verlagen, die allerdings im Verein nicht stimmberechtigt sind.

Das *Bildarchiv prometheus* wurde von 2001–2004 vom BMBF als Verbundprojekt gefördert. Träger ist seitdem ein eingetragener, gemeinnütziger Verein zur Förderung von Wissenschaft und Forschung mit Sitz am Kunsthistorischen Institut der Universität zu Köln, der die Verwertung organisiert. Mitglieder können Einzel- sowie juristische Personen sein. Der Mitgliedsbeitrag ist gestaffelt und beträgt auch für Institutionen unter 200 Euro pro Jahr. Bei beiden Diensten sind die Mitgliedsbeiträge nicht kostendeckend kalkuliert, es werden zusätzliche Einnahmen aus der Monetarisierung von Daten oder Leistungen erzielt (ohne Gewinnerzielungsabsicht). Im Fall von *prometheus* werden im Zuge der Verwertung individuelle, Campus- und Institutionenlizenzen erteilt, die jährlich zu erneuern sind. Die bildgebenden Einrichtungen stellen ihre Objekte für die Forschung und Lehre kostenlos zur Verfügung. Des Weiteren erhält der Verein eine Förderung von Stiftungen sowie Universitäten.

3.5.2 Monetarisierungsmodelle

Etwa die Hälfte der betrachteten Dienste nutzt in mehr oder weniger großem Umfang Monetarisierungsmodelle – nicht zur vermarktenden Gewinnerzielung, aber zur Refinanzierung des Angebots. Dazu gehören auch solche Dienste, die wissenschaftsintern organisiert sind – wobei die zu zahlenden Summen in der Regel dann als Deckungsbeitrag (Gebühren) ohne Gewinnerzielungsabsicht angelegt sind.

Gebühren nach dem Modell des Publikationssystems

Forschungsdatenrepositorien erheben **Gebühren** oder Entgelte typischerweise von den Datenerzeugern für die Nutzung der angebotenen technischen Infrastruktur (Supply-Side-Modell, vgl. Kap. 1.2). Die Gebühren dienen zur Deckung der Kosten für die Veröffentlichung, die Archivierung und ggf. Kuratierung. Sie

ähneln den Article Processing Charges (APC) im wissenschaftlichen Publikationssystem dahingehend, dass hier der Erzeuger des Inhalts für die Dienstleistung bezahlt. Kalkuliert wird jedoch nicht gewinnorientiert. Es finden sich zudem kostenlose Varianten, die von Einrichtungen im Rahmen ihres Auftrags bereitgestellt und von den Trägern finanziert werden.

Die Verwertungsplattformen für Satellitendaten aus dem europäischen COPERNICUS Programm bieten auch infrastrukturelle Services. Hier werden neben Datenzugang, Rechen- und Speicherkapazitäten auch Applikationen zur Datenanalyse monetarisiert, nicht die Daten selbst. Auch Dritte können auf diesen Plattformen Datenprodukte oder Applikationen vertreiben. Die Leistungen werden den Datennachfragern berechnet.

Monetarisierung
von Applikationen
zur Datenanalyse

Die Bezahlweise kann bei kommerziellen wie wissenschaftlichen Anbietern stark variieren und beispielsweise zeitlich begrenzte (erneuerbare) Lizenz- oder Subskriptionsgebühren umfassen. Auch finden sich sogenannte Pay-per-use-Modelle (Staffelungen nach Verbrauch bzw. Umfang der Nutzung, zum Beispiel Speichervolumen, Anzahl der Downloads pro Monat, Anzahl angeforderter Datenbankzeilen usw.) Für akademische Einrichtungen räumen kommerzielle Dienste teilweise Sonderkonditionen ein. Dazu gehören (Campus-)Lizenzen oder gesonderte Angebote für Studentinnen und Studenten. Der Leistungsumfang solcher verbilligten Angebote ist gegenüber dem regulären zahlungspflichtigen Angebot allerdings nicht selten reduziert. Andere Anbieter bemühen sich um Reduktion von Bezahlschranken für weniger solvente Klienten: *TAIR* verzichtet bei Datennachfragern aus Entwicklungs- bzw. Schwellenländern auf die Subskriptionsgebühren. Nach einem Jahr werden alle Inhalte offengestellt. *Dryad* listet mehr als ein Dutzend Sponsoren auf, die ggf. die Data Publishing Charge (DPC) für bedürftige Forscherinnen und Forscher übernehmen.

Gebührenstaffelungen
nach Verbrauch bzw.
Umfang der Nutzung

Als weitere Monetarisierungsstrategien finden sich die Lizenzierung von Technologien an andere Verwerter/Vermarkter und Sponsoring. Andere bieten zusätzlich zu Basis-Diensten gesonderte Verträge für eine Implementierung an der Einrichtung: Customizing-Optionen (Einbindung eigener Rechenzentren), institutionseigenes Präfix bei der DOI und Anpassung an das Corporate Design.⁷⁶

Monetarisierungsstrategien betreffen in vielen Fällen nicht den gesamten Dienst, sondern nur bestimmte Teile davon (Premium-Angebote), während Basisangebote – sozusagen zum Einstieg – kostenfrei sind. Für dieses Erlösmodell wird der Neologismus **Freemium-Modell** verwendet.

Freemium:
Basisangebote + X

⁷⁶ Soltau et al. (2019) – Management digitaler Forschungsdaten.

So kann bei offenem Basis-Zugang der Zugang zu qualitativ hochwertigen Inhalten kostenpflichtig sein, wie im Fall der *Deutschen Fotothek*, die Bereitstellungs-entgelte für hochauflösende Bilder und für kommerzielle Nutzung erhebt. In anderen Diensten ist es der Zeitpunkt des Zugangs – bei *TAIR* ist der Zugang im ersten Jahr den zahlenden Abonnenten vorbehalten, ganz ähnlich den Embargo-Fristen bei einigen Open Access-Modellen. Eine weitere Form kostenpflichtiger Mehrwerte bezieht sich auf den Funktionsumfang, zum Beispiel den Download als PDF oder die Nutzung von APIs, sowie auf ein Zusatzangebot an Services. Das generische Repositorium *RADAR* bietet zum Beispiel gegen Entgelt ein Meta-datenmapping an. Bei anderen Datenplattformen können Visualisierungstools oder andere Applikationen hinzugebucht werden.

Risikokapital-
finanzierung noch
ohne Bedeutung

Angebote aus der digitalen Start-up-Szene sind im Bereich der wissenschaftlich relevanten Datendienste noch die Ausnahme. Ein Beispiel für **Venture Capital-Finanzierung** ist allerdings *TriNetX* – eine Unternehmensgründung im Bereich werthaltiger Daten, in welchem besonders hohes wirtschaftliches Wachstum erwartet wird (klinische Forschung).

Crowdfunding-Modelle finden sich im Bereich des Open-Access-Publizierens. Dabei sammeln Verlage von teilnehmenden Bibliotheken Mittel ein, um ein bestimmtes Kontingent an Publikationen zu finanzieren, die der Verlag dann dauerhaft via Open Access zur Verfügung stellt. Unter den vom RfII betrachteten Datendiensten kamen derartige Modelle allerdings nicht vor. Dies ist wenig verwunderlich, denn bei jährlich zu erneuernden Verträgen würden hier hohe Transaktionskosten entstehen und die Planungssicherheit aber auch die Nachhaltigkeit für die dauerhafte Archivierung und Qualitätssicherung von Forschungsdaten könnte kaum garantiert werden.⁷⁷

⁷⁷ OECD (2017) – Business models for data repositories.

4 ZUSAMMENFASSUNG UND BEWERTUNG DER BEFUNDE

4.1 ZUSAMMENFASSUNG

Um den Status quo und Stellenwert heutiger in der Wissenschaft genutzter Datendienste zu verstehen, erscheint es wichtig, deren historische Genese zu verstehen: Sie entstehen aus den Forschungsformen und Forschungsparadigmen verschiedener wissenschaftlicher Domänen.

Um zu übergeordneten Vergleichsmaßstäben zu kommen, die Aufschluss über die Ausgestaltung von Betriebs- und Geschäftsmodellen geben, die mit wissenschaftlichen Bedarfen und Interessen konformgehen, hat die Fallbeispielanalyse domänenübergreifende Merkmale genutzt. So nähert man sich einer Perspektive, die das gesamte Wissenschaftssystem im Blick hat und Einschätzungen in der Frage erlaubt, welche Chancen und Risiken sich hinsichtlich der wissenschaftlichen Nutzung und dem wissenschaftskonformen Transfer von Daten ergeben, wenn Datendienste mit verwertungsorientierten Geschäftsmodellen arbeiten.

Vergleichsmaßstäbe:
Perspektive auf
das gesamte
Wissenschaftssystem

In den in der Fallbeispielanalyse behandelten Diensten sind insbesondere solche Merkmale und Ausprägungen identifiziert worden, die idealtypisch dem Ziel der weiteren Unterstützung wissenschaftlichen Arbeitens in allen Disziplinen und Domänen entgegenkommen. Unter diesem Dach übergreifender wissenschaftskonformer Merkmale lassen sich dann im Spannungsfeld von Nutzung oder aber (bzw. auch) Verwertung und Vermarktung einzelne Komponenten in den Betriebs- und Geschäftsmodellen gewichten – und zwar unterschiedlich, je nach Anwendungsbezug, Wirtschaftsnähe und bereits gelebten Routinen des Austauschs mit kommerziellen Datennachfragern (z.B. in der Chemie, den Lebenswissenschaften, den Ingenieurwissenschaften aber auch Bereichen wie den Rechtswissenschaften).

Welche Betriebs- und Geschäftsmodelle von Diensten sich als wissenschaftskonform erweisen, lässt sich nicht pauschal beantworten: Zu sehr hängt dies von den individuellen Bedürfnissen der Wissenschaftlerinnen und Wissenschaftler ab. Zusammenfassend lassen sich im Forschungsprozess domänenübergreifend folgende **pragmatische Bedarfe** für Leistungen von Datendiensten identifizieren:

Bedarfe für
Bewertung
ausschlaggebend

- möglichst zügige und unkomplizierte **Publikation** von Forschungsinformationen und deren Verknüpfung mit verwandten Themen – dies kann vom Teilen relevanter Forschungsdaten bis zur Ergebnis- und Datenpublikation reichen;
- langfristige **Sicherung** und **Archivierung** von Daten und Datensätzen;
- **Verfügbarkeit** von umfassenden Zusammenstellungen von Daten, Informationen und Wissen (Suchen und Finden in einem Marktplatzmodell), quantitative extensive Zugriffsmöglichkeiten auf Daten unterschiedlicher Provenienz und Qualität (Mindestanforderung: FAIR);

- Arbeitsmöglichkeiten mit qualitativ sehr hochwertigen, d.h. wissenschaftlich kuratierten Datensammlungen zu einzelnen Themen (FAIR + Q = „gelingender Zugriff“ über ein Redaktionsmodell, das heißt: über das bloße Suchen und Finden hinaus).

Pluriversum von Betriebs- und Geschäftsmodellen funktional

Eine einheitliche Lösung zeichnet sich weder für diese generalisierbaren Bedarfe von Wissenschaftlerinnen und Wissenschaftlern noch für domänenspezifische Datendienste ab. Bereits die Analyse der Entstehung von wissenschaftlichen Diensten oder die Dienstbarmachung nicht primär wissenschaftlicher Dienste für Forschungszwecke aus den Nutzungslogiken der Domänen heraus hat gezeigt, dass ein Pluriversum von Betriebs- und Geschäftsmodellen möglich und funktional sein kann – teilweise lassen sich auch Betriebs- aber keine Geschäftsmodelle identifizieren (s.u.). Dies lässt sich nicht zuletzt damit erklären, dass zahlreiche wissenschaftliche Datendienste mit der Hilfe von staatlichen Projektförderungen in der Wissenschaft selbst entstanden sind, wobei Fragen der (Re-)Finanzierung oder gar Monetarisierung der eigenen Angebote zunächst von untergeordneter Bedeutung waren. Zieht man die Analyse nach übergreifenden Merkmalen und deren Ausprägungen noch hinzu, wird das Bild noch einmal bunter:

Orientierung am Publikationssektor

Die Fallbeispiele zeigen eine Vielfalt an Geschäftsmodellen für Datendienste, die sich zunächst einmal nicht wesentlich vom Publikationssektor unterscheiden: Subskription, Varianten von „Contributor Pays“ oder Freemium-Angebote (= Mix aus kostenfreien Basisangeboten und hinzubuchbaren kostenpflichtigen Premiumangeboten, siehe Kap. 3). Gerade letzteres ist ein in der Digitalwirtschaft gängiges Modell und verdient Aufmerksamkeit: Es bietet wissenschaftlichen Einrichtungen die Möglichkeit, Daten grundsätzlich kostenfrei für eine wissenschaftliche oder gewerbliche Nutzung bereitzustellen, und kann über zusätzliche Angebote Kostendeckungsbeiträge oder Gewinne erwirtschaften. Dort, wo solche Angebote von kommerziellen Anbietern offeriert werden, werden auf kostenfreier Basis teilweise weniger attraktive oder nicht auf anspruchsvolle Weise nutzbare Daten angeboten werden. Über diesen Weg können auch zuvor inbegriffene Leistungen kostenpflichtig ausgekoppelt werden und so implizite Kostensteigerungen entstehen.

Keine profitorientierten Geschäftsmodelle im wissenschaftlichen Betrieb

Genuin im wissenschaftlichen Betrieb entwickelte Dienste sind häufig kostenfrei verfügbar, sofern keine arbeitsintensiven Sonderleistungen angefragt werden (z.B. hochauflösende Bilder oder Ähnliches). Auch für deren Erstellung werden aber lediglich moderate Gebühren erhoben. Mit diesen Diensten sind bislang keine expliziten Geschäftsmodelle verknüpft. Sie werden aus Projektmitteln (solange diese zur Verfügung stehen) oder aus den institutionellen Haushalten wissenschaftlicher Trägerorganisationen subventioniert und damit letztlich in beiden Fällen: aus Mitteln der staatlichen Wissenschaftsförderung.

Kommerzielle Dienste haben sich häufig dort entwickelt, wo heterogene oder quantitativ sehr umfangreiche Datenmengen verwertet werden. Viele dieser Dienste sind in ihren Basisleistungen zunächst kostenfrei. Eigene Wertschöpfung erzielen sie teils über Premiumangebote oder eventuell bei anderen Kundenkreisen (zweiseitige Märkte bei differenzierter Nachfrage, siehe Kap. 1). Ein Geschäftsfeld kommerzieller Anbieter sind zudem Infrastrukturdienstleistungen für Wissenschaftseinrichtungen oder Communities (Platform-as-a-Service, Infrastructure-as-a-Service, Software-as-a-Service). In üblichen Vergabeverfahren werden hier in der Regel Dienste bevorzugt, die von bereits länger und erfolgreich am Markt agierenden Anbietern offeriert werden. Dagegen haben kleine Start-up-Dienste oder Non-Profit-Initiativen, die Open Source-Lösungen entwickelt haben, mit Blick auf die Forderungskataloge öffentlicher Ausschreibungen auch in der Wissenschaft ggf. das Nachsehen. Aufgrund kleinerer Unternehmenskapitalisierung können sie Zusagen zur Dauerhaftigkeit und Nachhaltigkeit ihres Angebots weniger glaubhaft belegen.

Kommerzielle Dienste:
Wertschöpfung durch
zweiseitige Märkte

4.2 BEWERTUNG

Ob ein Dienst aus der übergeordneten Perspektive eines nachhaltigen Datenökosystems für die Wissenschaft besser öffentlich, in privater Trägerschaft (also: kommerziell) oder als Public-Privat-Partnership angeboten werden sollte, lässt sich, wie bereits festgehalten, nicht abstrakt entscheiden. Für eine adäquate Einschätzung kommt es auf den Nutzungszweck wie auch auf die Relevanz und konkrete Bewertung von Nachhaltigkeitsdimensionen an. Geht es darum, der Wissenschaft (und anderen Nutzern) den Zugriff auf ein möglichst breites und diversifiziertes Portfolio an Daten bereitzustellen, kann ein kommerzieller Betrieb effizient sein. Freilich um den Preis, dass man sich auf die Typik eines heterogenen „Marktplatzes“ einlassen muss und dass sich bei fachlich sehr heterogenen Diensten i.d.R. eine Qualitätskontrolle der dort verfügbaren Daten durch die wissenschaftlichen Nutzer nur mit hohem Aufwand oder gar nicht durchführen lässt.

Qualitäts- und
Zugangssicherung als
Bewertungskriterien

Entsprechend gibt es Überlegungen, im Rahmen fachübergreifender Dateninfrastrukturen wie z.B. der NFDI oder der EOSC selbst auf wissenschaftliche Nutzung ausgerichtete Archive und Plattformen zu entwickeln, die das bestehende, wissenschaftlich kuratierte Angebot ergänzen und verknüpfen können. Aus der Perspektive des einzelnen Wissenschaftlers sind Dienste, die Archiv-, Plattform- oder Datenbanktechnologien anbieten attraktiv, wenn sie das Suchen und Finden heterogener Daten und Forschungsinformationen erleichtern. Zumindest diese Funktion sollte bei Nutzung kommerzieller Dienste für Wissenschaftlerinnen und Wissenschaftler auch weiterhin kostenfrei möglich sein.

Mittel- bis langfristige Stabilität der Nutzungsbedingungen und der Kosten wichtig

Bei einer Nutzung der Angebote kommerzieller Dienste können Preise sowie Nutzungsbedingungen sich verändern, da die mittel- bis langfristige Stabilität des Dienstes an dessen Rentabilität gebunden ist. Hier liegen Risiken für die Wissenschaft: Wird der Betreiber insolvent oder stellt den Dienst ein, können Daten verloren gehen. Die kommerziellen Dienste, bei denen bevorzugt Long-Tail-Daten für „Supplement-Publikationen“ zu wissenschaftlichen Artikeln abgelegt werden, sind teilweise in das internationale, öffentlich-private Archivierungssystem für wissenschaftliche Literatur eingebunden, ein Anspruch auf Verfügbarkeit besteht jedoch nicht. Auch Datenrepositorien z.B. an Universitäten oder aus Projektförderungen entstandene Archive sind vor Einstellung nicht gefeit und können mangels weiterer Förderung und Pflege schlicht „eingehen“. Auf die organisatorischen Lücken im Bereich der Langzeitarchivierung hatte der Rfll in seinem Positionspapier LEISTUNG AUS VIELFALT bereits hingewiesen.

Einbindung der Wissenschaft in die Governance von Bedeutung

Dienste, bei denen über das Suchen und Finden hinaus bereits die Funktionalität einer qualitativ hochwertigen wissenschaftlichen Nutzung der Daten zu Forschungszwecken mitgedacht wird, sind häufig als „Redaktionsmodell“ organisiert. Das heißt, die Daten werden dort durch wissenschaftliches Personal nach wissenschaftlichen Kriterien kuratiert – was sie auch für kommerzielle Verwertungen jenseits reiner Big-Data-Analysen interessant macht. Die Analyse der beteiligten Akteure zeigt, dass hier in der Regel auch Wissenschaftlerinnen und Wissenschaftler als Nachfrager und wissenschaftliche Einrichtungen als Betreiber in die Steuerung eingebunden sind. Entsprechende Dienste, wie sie beispielsweise von Forschungsdatenzentren zur Verfügung gestellt werden, sind zumindest in Deutschland meistens in öffentliche Trägerschaften eingebettet und damit institutionell finanziert. Es fanden sich im Einzelfall auch eigens gegründete Betreibergesellschaften für solche Community-orientierten Dienste (non-profit). Moderate Gebührenmodelle dienen bestenfalls der Re-Finanzierung des Kuratierungsaufwands, nicht der Gewinnerzielung (was für Anfragen zur gewerblichen Verwertung aber sicherlich eine Option wäre – siehe das in Kap. 1 erwähnte Geschäftsmodell des „zweiseitigen Marktes“).

Die Zusammenarbeit mit Intermediären hat Potenzial

Die Fallbeispielanalyse zeigt eine Vielzahl wissenschaftlicher und kommerzieller Akteure, die als Intermediäre eine Brücke zwischen Datenerzeugern und Datennachfragern schlagen (siehe auch Kap.1). Mit der Hinzuziehung solcher Intermediäre für einen Dienst kann ein erhebliches Potenzial verbunden sein, was die Kapazitäten für die Entwicklung attraktiver Dienste bzw. Datenprodukte betrifft. Dies betrifft vor allem den Zugang zu einer größeren Breite und Diversität an Ressourcen und die zielgruppenspezifische Aufbereitung von Daten.

Gefahr der Abhängigkeit von Intermediären mit starker Marktstellung

Gleichwohl sind Abhängigkeiten und Asymmetrien naheliegend, wenn Intermediäre mit starker Marktstellung erhebliche Anteile wissenschaftlicher Information aggregieren und damit „unentbehrlich“ werden. Dies geschieht bei großen Datenkonzernen, durchaus aber auch bei wissenschaftsinternen Diensten, wie

der Fall der *Worldwide Protein Data Bank* zeigt: Als nach dreißig Jahren Geldgeber und Betreiber dieser Datenbank ausfielen, musste die wissenschaftliche Gemeinschaft eine Lösung für die Weiterführung finden. Die Ressource wird mittlerweile gemeinschaftlich, als eine Art Allmende, durch regionale, öffentlich finanzierte Datenzentren getragen. Im kommerziellen Bereich dürften Zusammenbrüche dieses Typs möglicherweise nicht auf diese Weise heilbar sein. Was bei einer Aufgabe des Dienstes mit den Daten geschehen würde, ist weitgehend unklar. Falls es dazu kommen sollte, wäre wohl nach einer gemeinschaftlichen Lösung zu suchen, um den Verlust der Ressourcen zu verhindern und die Funktionalität von Diensten, beispielweise für die Nachweissysteme im Bibliothekswesen, nicht zu beeinträchtigen.

Erheblicher als ein Ausfallszenario mag noch das Risiko künftiger Preissteigerungen sein. Sollte sich ein Dienst durch die Aggregation großer Datenmengen im Lauf der Zeit unentbehrlich machen bzw. „systemrelevant“ werden, kann er die Preise nach oben treiben, ohne dass die Wissenschaft dem etwas entgegensetzen könnte.

Eine **(robuste) Pluralität der Angebote** und eine wissenschaftsinterne Absicherung der Datenbestände, die bei kommerziellen Dienstleistern vorgehalten werden, erscheint daher geboten. Beispielgebend könnte hierfür die im Rahmen der Allianz-Initiative „Digitale Informationen“ entwickelte und seit 2014 mittels DFG-Projektförderung auf den Weg gebrachte Strategie zum Nationalen Hosting sein. Um „Störungen“ des wissenschaftlichen Betriebs durch prekären Datenzugriff z.B. im Rahmen des Transfers oder der Einstellung von Fachzeitschriften, Abbestellung elektronischer Ressourcen, Serverausfällen bei Verlagen etc. zu vermeiden, werden wissenschaftliche Informationen im Nationalen Hosting durch einen Archivdienst und perspektivisch durch ein weiteres digitales Netzwerk abgesichert, so gut dies im Einzelfall möglich ist.⁷⁸

Aus Perspektive des Wissenschaftssystems ist es hochgradig bedeutsam, Forscherinnen und Forscher als Akteure in die Ausgestaltung und Weiterentwicklung von Datendiensten einzubinden. Diese können dort in ganz unterschiedlichen Rollen auftreten, die vielfach auch konvergieren: am augenscheinlichsten in der Doppelrolle als Datenerzeuger und Datennachfrager, zusätzlich als Anbieter, Betreiber, Datensubjekt, Kurator etc. Als wissenschaftsgerecht wird der Umgang mit Forschungsdaten in all diesen Rollenkonstellationen nur dann erfahren, wenn er der Dynamik von Forschungsprozessen und Veränderungen in der Theorie- und Methodenentwicklung von Disziplinen, interdisziplinären Forschungsfeldern und Domänen folgt. Der RfII hat in seinen Empfehlungen

Kostenunsicherheit
und Ausfallszenarien
als Bedrohung

Wissenschaftsinterne
Absicherung der
Datenbestände nötig

Einbindung der
Forschung für die
Weiterentwicklung
von Datendiensten
essentiell

⁷⁸ nathosting.de/display/ND/Hintergrund (zuletzt abgerufen am 19.05.2021).

zu Forschungsdatendiensten wiederholt auf die Bedeutung der Mitwirkung der Nutzerinnen und Nutzer von Diensten hingewiesen. Wie die Auswahl, Kuratierung und technische Umsetzung der Inhalte gesteuert wird bzw. wer über den Zugang und ggf. Ausschluss bestimmter Akteure entscheidet, ist für die wissenschaftliche Qualität von Datendiensten von großer Bedeutung. Und: Vertrauen in diese Qualität ist auch für die gewerbliche Datennachfrage – also das Vermarktungspotenzial der Daten – relevant, denn genau darin liegt auch in der kommerziellen Verwertung der Mehrwert wissenschaftlicher Daten.

Prinzip der besten Eignung

Bei der Frage, ob wissenschaftskonforme Datendienste kommerziell am Markt oder in öffentlicher Trägerschaft angeboten werden sollten, ist letztlich das Prinzip der besten Eignung⁷⁹ ausschlaggebend. Die Fallbeispielanalyse gibt keinen direkten Hinweis darauf, dass Grundsatzentscheidungen klar zugunsten einer der beiden Seiten getroffen werden müssten. Hier scheinen im Gegenteil Mischformen vielversprechend zu sein, in denen öffentlich geförderte Angebote mit kommerziellen Angeboten verzahnt werden können. Partizipation von Fachgemeinschaften und öffentliche Aufmerksamkeit können den Bestand solcher Arrangements auch „politisch“ sichern.

Sicherstellung eines nachhaltigen Basisangebots durch die öffentliche Hand nötig

Zentral für ein Gelingen der Auswahl nach bester Eignung bleibt die Verantwortung der öffentlichen Hand für die Sicherstellung eines dauerhaften und nachhaltigen Basisangebots an die Wissenschaft, dass mindestens die Funktionalitäten des Suchens und Findens sowie des Bewahrens (Langzeitspeicherung auf maschinenlesbaren Medien) abdeckt. Weiterführende Funktionalitäten der qualitätsgesicherten Publikation und forschenden Nachnutzung von Forschungsdaten können sowohl in wissenschaftlicher Selbstorganisation und mit Mitteln der institutionellen Förderung (für die Wissenschaft oder einzelne Communities „systemrelevante Dienste“) und Projektförderung (Innovationen) als auch mit kommerziellen Partnern (Archiv- und Plattformdienste, Innovationen) realisiert werden. Bei Mandatierung kommerzieller Anbieter müssen die Vertragsbedingungen jedoch dringend wissenschaftskonform ausgestaltet werden.

⁷⁹ RfII (2020) – Datendienste nachhaltig gestalten. Das vom RfII formulierte Prinzip besagt, dass die Aufgabenteilungen zwischen den innerwissenschaftlichen Verbundpartnern als auch mit externen, wissenschaftlichen sowie kommerziellen Akteuren auszuhandeln sind, „Handlungsleitendes Prinzip sollte dabei – nebst der Wahrung der Souveränität der Wissenschaft – die Übertragung einer Aufgabe an den jeweils fachlich-technisch am besten geeigneten Akteur sein. Grundlage ist eine klare Beschreibung der Aufgaben und ein Selbstverständnis darüber, welche Verantwortlichkeiten in der Hand der Wissenschaft verbleiben und welche Aufgaben in Zusammenarbeit mit externen Dienstleistern oder Partnern gut, international anschlussfähig und vor allem wissenschaftskonform gelöst werden können.“

5 EMPFEHLUNGEN

Auf Basis der analysierten Fallkonstellationen (Kapitel 2 und 3) und ihrer Bewertung (Kapitel 4) gibt der RfII abschließend erste Empfehlungen für die Ausgestaltung von Regulierungsanstrengungen ab. Diese richten sich an Bund und Länder, Wissenschaftseinrichtungen, Hochschulen sowie die wissenschaftlichen Fachgemeinschaften, Domänen und die hier tätigen Forscherinnen und Forscher.

Tabelle 4: Empfehlungen gruppiert nach Adressaten

Zur staatlichen Regulierung des Marktes für kommerzielle und öffentlich finanzierte Datendienste
5.1 Selbstbestimmtes Teilen von Forschungsdaten ermöglichen und innerwissenschaftliche Freiräume schützen
5.2 Hürden für den Marktzugang kleiner Anbieter und innovativer Dienste senken
5.3 Datentreuhandmodelle wissenschaftsfreundlich ausgestalten und fördern – Verfahren der Qualitätssicherung aus der Wissenschaft maßvoll adaptieren
Zur Gestaltung der Kooperationsbeziehungen zwischen wissenschaftlichen Akteuren und externen Anbietern von Diensten
5.4 Wissenschaftlichen Zugriff auch bei Monetarisierung von Daten und datenbezogenen Leistungen sichern
5.5 Datenverluste vermeiden – dauerhaften Zugang zu Daten und Code vertraglich sichern
5.6 Wissenschaftliche Qualitätssicherung gewährleisten – auch auf Seiten des Kooperationspartners
5.7 Selektions- und Bewertungskriterien für Forschungsdaten anhand wissenschaftlicher Vorgaben entwickeln
5.8 Fachlich-inhaltliche Aufbereitung der Daten für außerwissenschaftliche Kontexte leisten
Zur (Selbst-)Organisation von Handlungsfähigkeit auf Seiten der Wissenschaft
5.9 Beratung und Vermittlung für Fachgemeinschaften organisieren – eine Chance für die NFDI und ihre Konsortien
5.10 Essentielle Datenbestände der wissenschaftlichen Fachgemeinschaften und Domänen definieren und sichern

Richtschnur für die hier präsentierten Empfehlungen ist: Im Bereich der Verwendung von Forschungsdaten sollte es möglichst weitreichend zu nachhaltigen Lösungen kommen. Angesprochen wird hierbei insbesondere die Kooperation mit kommerziellen Anbietern von Diensten. Ebenso bedeutsam ist aber der Umgang der Wissenschaft mit wirtschaftlich ausgerichteten Handlungsmotiven, die auch bei Diensten auftreten können, die in der Wissenschaft selbst aufgebaut wurden.

Solche Motive könnten bei einem möglichen Rechtsanspruch Dritter auf unlimitierten Zugang zu Daten aus der öffentlich finanzierten Forschung noch gesteigert werden. Hier müssen ebenfalls Nutzung und Verwertung im Einklang mit den für das Wissenschaftssystem essentiellen Zielstellungen und Bedarfen stehen.⁸⁰

Primat der
Forschungsmethodik
ist nicht verhandelbar

Der Rfll hat hier insbesondere Fallkonstellationen im Blick, in denen der auf Wissenschaftsseite nicht verhandelbare Primat der Forschungsmethodik, der wissenschaftliche Wettbewerb selbst (der im Kern kein wirtschaftlicher, sondern ein auf Urheberschaft gegründeter reputationsorientierter bleibt) oder aber das Erhaltungsinteresse an essentiellen Forschungsdatenbeständen (als – ggf. auch zunächst – *nur* für die Wissenschaft gemeinfreie Ressource) in ein Spannungsverhältnis zu unter Umständen ganz anders gearteten Verwertungs- oder Vermarktungsinteressen geraten kann. Wissenschaftliche Akteure müssen auch in Kooperation mit kommerziell arbeitenden Partnern die Verfügungs- und Definitionshoheit („Souveränität“) in Fragen behalten können, die

- a) sowohl die Güte und qualitative Bewertung von genuin durch Forschung erzeugte Daten betreffen als auch
- b) die Beurteilung der qualitativen Eignung außerhalb von Forschungskontexten bereitgestellter Daten für Zwecke wissenschaftlicher Nutzung betreffen (wie das z.B. in den Forschungsdatenzentren von Ämtern und Behörden geschieht).

Souveränität
der (digitalen)
Wissenschaft

Eine Souveränität der (digitalen) Wissenschaft in diesem Sinne kann nach Auffassung des Rfll nur erlangt bzw. gesichert werden, wenn die europäischen und nationalstaatlichen politischen Institutionen durch ihre umfangreichen aktuellen Vorhaben zur Regulierung der digitalen Märkte Rahmenbedingungen schaffen, die die Wissenschaft dazu ertüchtigen, mit kommerziellen Anbietern – auch mit multinationalen und außereuropäischen Unternehmen, die digitale Dienstleistungen für die Wissenschaft anbieten – „auf Augenhöhe“ zu verhandeln. Gleichzeitig ist ein bestimmter, in Deutschland auch in der Verfassung festgeschriebener Auftrag zu erfüllen: Wissenschaftliche Einrichtungen und intermediäre Institutionen, die mit außerwissenschaftlichen Anbietern kooperieren, die eventuell hierfür sogar eigens mandatiert sind, müssen die Bedürfnisse von Forschung und Lehre insgesamt und langfristig im Auge behalten.

⁸⁰ Das vorliegende Positionspapier betrachtet den Themenkomplex der Nutzung und Verwertung von Forschungsdaten nicht nur „in“ Deutschland, aber aus der Perspektive des deutschen Wissenschaftssystems. Vorgaben für die Bereitstellung und das Teilen von Forschungsdaten durch nationale oder supranationale Gesetzgebungen können tiefgreifende Konsequenzen für internationale Forschungsk Kooperationen im globalen Wissenschaftssystem haben und erzeugen in multinationalen Forschungsverbünde schon heute Kooperationshindernisse, da nicht hinreichend deutlich ist, welche und unter welchen Konditionen von einzelnen Partnern erhobene Daten geteilt und weitergegeben werden dürfen. Zu dieser Problematik bedarf es in Zukunft einer gesonderten Betrachtung.

Geeignete Kontrollinstrumente und hohe Transparenz können dazu beitragen, dies auch mittelfristig zu garantieren. Des Weiteren müssen wissenschaftlichen Einrichtungen oder Intermediäre durch einen klugen Ordnungsrahmen in die Lage versetzt werden, ihre Belange gegenüber solchen Akteuren ein- und durchsetzen zu können, die mit erheblicher Marktmacht ausgestattet sind. Dies schließt öffentlich garantierte Rückfalloptionen ein, Daten und Datendienste notfalls auch in wissenschaftlicher Selbstorganisation und mit Hilfe öffentlicher Förderung zu organisieren, um Forschungsdaten auch dann weiterverarbeiten und teilen zu können, wenn dies für kommerzielle Anbieter nicht bzw. nicht mehr rentabel ist. Hierbei können sich für unterschiedliche Forschungsfelder und Domänen auch verschiedene Lösungsansätze anbieten: vollständige Eigenregie bzw. der Aufbau innerwissenschaftlicher Dienste, Auslagerung an kommerzielle Anbieter oder Aufbau hybrider Organisationen im Rahmen von Public-Private-Partnerships.

Der RfII versteht die Fallbeispielanalyse und die auf ihrer Basis erfolgte Abwägung zum Einsatz kommerzieller und nicht-kommerzieller bzw. genuin wissenschaftlicher Intermediäre auch als einen Beitrag zur Wiedergewinnung und letztlich zur Steigerung der Souveränität der Wissenschaft in einer zunehmend digitalen Welt. Die folgenden Empfehlungen sind insofern auch als empirisch fundierte Antworten auf eine Diagnose zu verstehen, die der Wissenschaftsrat in seinem Positionspapier *IMPULSE FÜR DIE COVID-19-KRISE FÜR DIE WEITERENTWICKLUNG DES WISSENSCHAFTSSYSTEMS IN DEUTSCHLAND* gestellt hat.⁸¹ Dort wird ausgeführt, dass unter den Gegebenheiten der globalen Pandemie in Forschung und Lehre die „Abhängigkeit von privatwirtschaftlichen Anbietern augenfällig geworden [sei ...]. Die Anbieter solcher Dienste und Plattformen fungieren mittlerweile als De-facto-Regulierer“.⁸² Um für die Wissenschaft ebenso schädlichen wie dauerhaften Lock-In-Effekten vorzubeugen, plädiert der Wissenschaftsrat dafür, dass das Wissenschaftssystem nicht nur „seine Souveränität im digitalen Raum ausbaut“,⁸³ sondern seine maßgeblichen Akteure auch eine „Vorreiterrolle für die souveräne Gestaltung des digitalen Raums und die Entwicklung seiner Potenziale“⁸⁴ übernehmen.

Ausdruck von Souveränität der Wissenschaft in dem hier verstandenen Sinne ist auch die Diversität der praktizierten Forschungsansätze, Theorien und Methoden sowie die Möglichkeit, dieser Vielfalt Ausdruck zu verleihen. Auch unter diesem Gesichtspunkt befürwortet der RfII Wettbewerb und Pluralität unter den Diensteanbietern und ihren jeweiligen Geschäftsmodellen. Um dies zu gewährleisten, adressiert der RfII Empfehlungen sowohl an die Ebene der staatlichen

Problemstellung:
Diensteanbieter als
„De-facto-Regulierer“
für die Wissenschaft

**Wettbewerb und
Pluralität als
Grundlage für
wissenschaftliche
Souveränität**

⁸¹ WR (2021) – Impulse aus der COVID-19-Krise.

⁸² Ebd. S. 44

⁸³ Ebd. S. 46

⁸⁴ Ebd. S. 48

Regulierungsakteure als auch an die wissenschaftlichen Einrichtungen, Hochschulen, die Fachgemeinschaften und die einzelnen Wissenschaftlerinnen und Wissenschaftler, die in Forschungsprozesse und damit einhergehend in unterschiedliche Stadien des forschungsbezogenen Datenlebenszyklus involviert sind. Nicht zuletzt sieht der RfII die derzeit im Aufbau befindliche Nationale Forschungsdateninfrastruktur (NFDI) und ihre Konsortien im Bereich der Nutzung und Verwertung von Forschungsdaten in einer Schlüsselposition, wenn es um die Vermittlung zwischen der Rationalität der wissenschaftlichen und der ökonomischen Sphäre geht.

EMPFEHLUNGEN ZUR STAATLICHEN REGULIERUNG DES MARKTES FÜR KOMMERZIELLE UND ÖFFENTLICH FINANZIERTER DATENDIENSTE

5.1 SELBSTBESTIMMTES TEILEN VON FORSCHUNGSDATEN ERMÖGLICHEN UND INNERWISSENSCHAFTLICHE FREIRÄUME SCHÜTZEN

Empfehlungen 5.1

- Frei- und Schutzräume für innerwissenschaftliche Nutzung und komplikationsloses Teilen von Daten definieren
- Verwendung von Forschungsdaten erst nach Einwilligung der datenerzeugenden Akteure gestatten
- Fragile Voraussetzungen der Autonomie von Forschung in der Gesetzgebung berücksichtigen

Grundlage für eine souveräne Gestaltung des globalen Forschungsraums durch seine maßgeblichen Akteure und Institutionen sind nach Auffassung des RfII politische Rahmenbedingungen, die ein komplikationsloses Teilen von qualitätsgeprüften Daten unter Wissenschaftlerinnen und Wissenschaftlern einerseits ermöglichen und andererseits hierbei die Eigenzeit von Forschungsprozessen berücksichtigen. Diese Rahmenbedingungen sind der Frage nach der Auswahl des jeweils angemessenen Datendienstes und seines Geschäfts- und Betriebsmodells noch vorgelagert – sie sind funktionale Voraussetzungen für wissenschaftlichen Erkenntnisgewinn.

Schutzräume für
kollegiales Datenteilen
sichern

Zu diesen unverzichtbaren Voraussetzungen gehören auch innerwissenschaftliche Üblichkeiten des Umgangs mit Daten.⁸⁵ Viele sind implizit und so selbstverständlich stets geschützt gewesen, dass Wissenschaftler darauf bisher fraglos vertrauen. Hierzu zählt, dass aktuell erhobene Daten und Datenspuren während laufender Forschungsprozesse noch nicht allgemein zugänglich sind, sondern die Forscherinnen und Forscher Zugriffe zunächst im Kreis ihrer wissenschaftlichen Kooperationspartner (auch aus der Industrieforschung) bzw. in ihren

kollegialen Netzwerken, innerhalb eines Faches, eines Forschungsfeldes oder einer Einrichtung ermöglichen (Siehe Abbildung 1, S. 11: kollegiales Teilen). Je nach Fachkultur und Forschungsfeld sind innerhalb der wissenschaftlichen Fachgemeinschaften unterschiedliche Vorgehensweisen üblich und auch geboten, bevor gewonnene Daten zum Zweck der innerwissenschaftlichen Arbeit zirkulieren. Die Vermarktung von Forschungsdaten durch wirtschaftliche Akteure ist vor diesem Hintergrund ein eigenständiger Schritt, der keinesfalls den Verlust der Daten für weitere wissenschaftliche Forschungen (Nutzung) oder für gemeinwohldienliche Anwendungen (Verwertung) nach sich ziehen darf. Darüber hinaus sind der Vermarktung im Rahmen des Vertrauensschutzes für sensible und personenbeziehbaren Daten auch handfeste Grenzen gesetzt – etwa durch Einwilligungserklärungen seitens der Datengebenden (etwa Teilnehmer an Befragungen und Interviews, Probandinnen und Patienten).

Der Rfll spricht sich dafür aus, dass das Interesse Dritter an Zugang zu Forschungsdaten im Forschungsprozess bzw. vor Abschluss einer jeweiligen wissenschaftlichen Studie auch weiterhin die Regeln und Regulierungen von wissenschaftlichen Fachgemeinschaften und Professionszusammenschlüssen zu respektieren hat. Forscherinnen und Forscher geben ihre Daten in der Regel durch Ergebnis- sowie Datenpublikationen für Zwecke der innerwissenschaftlichen (Nach-)Nutzung und Verwertung frei. Freigaben zu Vermarktungszwecken können erfolgen, wenn dadurch

- die weitere Verfolgung der mit der Datenerhebung verbundenen wissenschaftlichen Zielsetzungen auch durch andere Forschende sowie
- eine gemeinwohldienliche Verwertung der Forschungsdaten durch andere öffentliche Stellen und Akteure der Zivilgesellschaft möglich bleibt.

Der Rfll bekräftigt den Grundsatz, dass mit öffentlichen Mitteln erzeugte wissenschaftliche Daten auch öffentlich zugänglich bleiben müssen. In diesem Zusammenhang wird das Teilen von Daten mit außerwissenschaftlichen Interessenten vom Rfll ausdrücklich begrüßt und kann beispielsweise in Public-Private-Partnerships – unter Wahrung der Forschungsautonomie – auf geeignete vertragliche Grundlagen gestellt werden. Eine bereits früh im Forschungsprozess greifende „Pflicht“ zur Herausgabe von Daten zur Vermarktung würde nach Auffassung des Rfll den freien, kollegialen Austausch von Informationen

⁸⁵ Der vom Rfll 2016 bestimmte und 2017 überarbeitete Forschungsdatenbegriff ist sehr breit und umfasst „jegliche Daten, die im Zuge wissenschaftlichen Arbeitens entstehen, z.B. durch Beobachtungen, Experimente, Simulationsrechnungen, Erhebungen, Befragungen, Quellenforschungen, Aufzeichnungen, Digitalisierung, Auswertungen. Zu Forschungsdaten werden auch solche, nicht selbst gewonnenen Daten gerechnet, auf die die Wissenschaft zu Forschungszwecken zugreift, um sie für den konkreten Forschungsprozess als methodisch erforderliche Grundlage zu nutzen“. Rfll (2016) – Leistung aus Vielfalt, S. A-13, überarbeitete Fassung 2017: rfll.de/de/themen/.

in der Wissenschaft dagegen empfindlich stören – bis hin zu ernsthaften Gefährdungen der wissenschaftlichen Souveränität. Die Folge wäre dann absehbar nicht mehr, sondern weniger Offenheit im Wissenschaftssystem.

Autonomie der Forschung wahren

Der Rfll empfiehlt daher den politischen Akteuren, bei den laufenden Gesetzgebungsdiskussionen zur Förderung von Innovation durch Rechtsansprüche auf Offenheit und Zugangschancen, die Autonomie der Forschung und deren fragile Voraussetzungen zu beachten. Frei- und Schutzräume für innerwissenschaftliches Datenteilen müssen im Rahmen aktueller und künftiger datenbezogener Rechtsetzungsprozesse darüber hinaus klar formuliert sein. Eine diesbezüglich zu fragmentierte und uneindeutige Rechtslage würde auf die Bereitschaft, Daten zu veröffentlichen, hemmend wirken.

5.2 HÜRDEN FÜR DEN MARKTZUGANG KLEINER ANBIETER UND INNOVATIVER DIENSTE SENKEN

Empfehlungen 5.2

- Durch Rahmengestaltung Lock-In-Effekte vermeiden und unkontrollierten Preissteigerungen vorbeugen
- Über Senkung rechtlicher und faktischer Hürden in öffentlichen Vergabeverfahren gleiche Wettbewerbschancen für wissenschaftsnahe KMU, Start-ups und von Non-Profit-Organisationen ermöglichen
- Versicherungslösungen für Datendienste ermöglichen, die Haftungsrisiken auffangen und die Sicherheit wissenschaftlicher Partner bei Leistungsausfall des Dienstes erhöhen

Rahmengestaltung verbessern – mehr Wettbewerb ermöglichen

Kommerzielle Lock-In-Effekte, die zu Oligopolen oder Quasi-Oligopolen privater Anbieter führen und mittel- bis langfristig durch die zunehmende Abhängigkeit von deren Angeboten zu unkontrollierbaren Preissteigerungen führen können, müssen nach Auffassung des Rfll durch geeignete wissenschaftspolitische und vertragliche Rahmengestaltung vermieden werden. In öffentlichen Vergabeverfahren sollten rechtliche und faktische Hürden so weit abgesenkt werden, dass wissenschaftsnahe KMU, Start-ups und Non-Profit-Organisationen im Vergleich zur heutigen Situation einen leichteren Marktzugang erhalten. Insbesondere Anbieter, die mit Open-Source-Lösungen arbeiten, haben es gegenwärtig schwer, die in öffentlichen Ausschreibungen verlangten Haftungsanforderungen zu erfüllen. Die im Rahmen des Vorschlags für einen European Data Governance Act (DGA) geführte Diskussion um Versicherungsangebote für Data Service Provider, die das individuelle Haftungsrisiko eines Anbieters begrenzen und zugleich öffentlichen Nachfragern die nötige Sicherheit bei Leistungsausfall oder Fahrlässigkeit des Anbieters bieten, eröffnen Handlungschancen, um auch für Dienstleistungen und Serviceprodukte kleiner innovativer Unternehmen die

Wettbewerbschancen zu verbessern.⁸⁶ Der RfII setzt sich deshalb nachdrücklich dafür ein, Versicherungslösungen – abgestuft je nach Qualitäts-/Schutzniveau und „Wert“ der Daten – auf der supranationalen Ebene und in der nationalstaatlichen Umsetzung zu fördern und ihren Einsatz auch für den Bereich wissenschaftlicher oder wissenschaftsnaher Dienste zu prüfen.

5.3 DATENTREUHANDMODELLE WISSENSCHAFTSFREUNDLICH AUSGESTALTEN UND FÖRDERN – VERFAHREN DER QUALITÄTSSICHERUNG AUS DER WISSENSCHAFT MAßVOLL ADAPTIEREN

Empfehlungen 5.3

- Forschungsklauseln gesetzlich verankern und Forschungsdatenzentren ausbauen
- Zertifizierungs- bzw. Akkreditierungsniveaus anhand der Güte- und Schutzbestimmungen der Daten bestimmen
- Datengebende wissenschaftliche Einrichtungen bei Aufnahme, Klassifizierung und Aufbereitung der Daten einbeziehen

Zu den Zielen sowohl des europäischen DGA-Vorschlags als auch der Datenstrategie der Bundesregierung gehört es, die allgemeine Bereitschaft zur Bereitstellung von Daten durch Individuen, Unternehmen und öffentliche Einrichtungen zu erhöhen und deren verantwortungsvolle Nutzung zu fördern. In diesem Zusammenhang begrüßt der RfII die Absichtserklärung der Bundesregierung, bestehende Forschungsdatenzentren (FDZ) bei Ämtern und Ressortforschungseinrichtungen weiter auszubauen und gegebenenfalls neue FDZ einzurichten. Um das in den Daten liegende Innovationspotenzial bestmöglich ausschöpfen zu können, sollen aber auch andere Dateninfrastrukturen weiter ausgebaut und weitere Anreize zum freiwilligen Datenteilen durch neue Kooperationsformen wie Datentreuhänder gesetzt werden.

Bereitschaft zur Datenbereitstellung steigern

Der RfII bestärkt die Bundesregierung in ihrer Absicht, bei neuen Gesetzgebungsvorhaben zu prüfen, in welchem Umfang forschungsfreundliche, barrierefreie Zugangsregeln (sog. Forschungsklauseln) für die unabhängige wissenschaftliche Forschung geschaffen werden können. Im aktuellen Ausarbeitungsstand des Entwurfs für einen europäischen DGA sind Forschungsklauseln noch ein Desiderat. Der RfII empfiehlt dringend die Berücksichtigung eines privilegierten Datenzugangs für die öffentlich oder gemeinnützig geförderte Forschung auch bei europäischen Rechtsetzungsakten, insbesondere im DGA sowie dem Digital

Forschungsklauseln gesetzlich verankern

⁸⁶ Vgl. RfII (2021) – Stellungnahme zum DGA.

Services Act, dem Digital Markets Act sowie einem künftigen Data Act, um den europäischen Innovationsraum zu stärken.⁸⁷

Datentreuhänder
als neue Form
intermediärer
Datenmittler
aufbauen

Sowohl der DGA als auch die Datenstrategie der Bundesregierung lassen noch weitgehend offen, wie das Konzept des Datentreuhänders inhaltlich zu füllen wäre. Obwohl die Wissenschaft über bereits etablierte eigene Formen des Datenteilens und über institutionalisierte FDZ verfügt, ist sie an der Ausgestaltung der Verfahren für das Teilen zivilgesellschaftlicher und ökonomischer Daten durch Datentreuhänder in hohem Maße interessiert und kann hierfür insbesondere in Fragen der Qualitätssicherung Best Practice-Beispiele beisteuern. Der Rfll empfiehlt vor diesem Hintergrund, Datentreuhänder als eine neue Form von intermediären Datenmittlern aufzubauen, die dezidiert die Qualität der von ihnen gehaltenen und vermittelten Daten sowie der hierauf basierenden Vermittlungsdienste und Sicherheitsstandards in gestuften Verfahren der Zertifizierung bzw. Akkreditierung nachweisen sollten. Erst Qualität kann wirklich Vertrauenswürdigkeit schaffen. Dabei stellt der Rfll durchaus in Rechnung, dass Datentreuhänder gemäß des (im zweiten Quartal 2021) aktuellen Entwurfs zu einem europäischen DGA gerade nicht selbst die inhaltliche „Veredelung“ von Daten zu ihrem Geschäftsmodell erheben dürfen (um ihre Neutralität als Vermittler zu wahren). Datentreuhänder müssen aber in der Lage sein, Daten nachhaltig aufzubewahren und standardisierte Beschreibungen der zu vermittelnden Datensätze, etwa Auskünfte zum Qualitätsniveau, zu erstellen, zu aktualisieren und zu kommunizieren. Geregelt Verfahren zur Sicherstellung dieser Leistungen müssen überprüfbar sein.

Zertifizierungs- bzw.
Akkreditierungs-
niveaus bestimmen

Entsprechende Zertifizierungs- bzw. Akkreditierungsniveaus lassen sich dann anhand der Güte und Schutzbedürftigkeit der Daten sowie der Verwendungszwecke bemessen, auf denen ein Datentreuhänder sein Geschäftsmodell aufbaut. Eine bloße Registrierung, wie bislang im Vorschlag zum europäischen DGA vorgesehen, erachtet der Rfll als nicht weitreichend genug. So erhalten weder potentielle Datengeber noch Datennutzer Aufschluss über das Qualitätsniveau der Dienste, denen sie ihre Daten anvertrauen sollen bzw. deren Datenpools sie nutzen oder verwerten wollen.

Datengebende
wissenschaftliche
Einrichtungen ein-
beziehen

Der Rfll weist nachdrücklich darauf hin, dass auch Daten, die nicht für Forschungszwecke Verwendung finden, nur dann einen Mehrwert für Innovationsketten in Wirtschaft und Gesellschaft haben, wenn sie mindestens im Rahmen ihrer Metadaten Auskunft zu ihrer Provenienz und anderen Kontexten ihrer Entstehung bzw. „Herstellung“ und ihres bisherigen Transfers bieten. Datengebende wissenschaftliche Einrichtungen sollten deshalb von Datentreuhändern in Fragen der Aufnahme, Klassifizierung und Aufbereitung der Daten einbezogen werden.

⁸⁷ Vgl. ebd.

5.4 WISSENSCHAFTLICHEN ZUGRIFF AUCH BEI MONETARISIERUNG VON DATEN UND DATENBEZOGENEN LEISTUNGEN SICHERN

Empfehlungen 5.4

- Kostenfreie wissenschaftliche Nutzung der Daten bei Übergabe an vermarktende Partner vertraglich sichern
- Dauerhaften und vollständigen Zugang und Zugriff vertraglich festlegen
- Gebührensysteme für die mit Datenübergabe eventuell verbundenen Dienstleistungen maßvoll ausgestalten und disziplinäre Besonderheiten berücksichtigen

Wo Daten oder datenbezogene Leistungen aus der Wissenschaft monetarisiert werden, empfiehlt der RfII, Verträge mit kommerziellen Kooperationspartnern so auszugestalten, dass Garantien für eine kostenfreie wissenschaftliche Nutzung sowie weitere, für die Wissenschaft wichtige Leistungen (dennoch) vorgesehen sind. Für Zwecke der Verwertung können ggf. moderate Kostenmodelle vereinbart werden. Bei jeder Abgabe von Daten an externe Kooperationspartner ist darüber hinaus sicherzustellen, dass der Zugriff auf diese Daten der Wissenschaft nicht vollständig oder dauerhaft entzogen wird. Mit Industriepartnern müssen deshalb Optionen für eine weitere wissenschaftliche Nutzung der Daten vereinbart werden. Dabei müssen auch disziplinäre Besonderheiten der Erzeugung, Prozessierung und Nutzung der Daten berücksichtigt werden, etwa hinsichtlich der Dauer der Aufbewahrung.

Kostenfreie wissenschaftliche Nutzung/
moderate Kostenmodelle bei Verwertung

5.5 DATENVERLUSTE VERMEIDEN – DAUERHAFTEN ZUGANG ZU DATEN UND CODE VERTRAGLICH SICHERN

Empfehlungen 5.5

- Nachhaltigkeit der Datenarchivierung und des Zugangs zu Daten auch im Falle des Verkaufs oder der Insolvenz des Diensteanbieters sichern
- Regeln für die Migrierbarkeit der Daten bei Vertragsende oder Einstellung des Dienstes festlegen
- Proprietäre Quellcodes insbesondere von Speicher- und Archivierungssoftware bei Treuhändern hinterlegen

Forschungsdaten dürfen nicht verloren gehen

Forschungsakteure müssen auch bei Einschaltung oder Nutzung kommerzieller Anbieter von Datendiensten dauerhaften Zugang zu den Daten behalten können, die sie einspeisen. Deshalb dürfen die Nachhaltigkeit der Datenarchivierung und der Zugang zu Daten auch bei Verkauf oder Insolvenz eines kommerziellen Dienstleisters/Partners oder bei der Einstellung unprofitabler Dienste nicht verloren gehen. Der RfII empfiehlt deshalb individuellen Forscherinnen und Forschern, wissenschaftlichen Einrichtungen sowie NFDI-Konsortien, die für ihre Domänen und Communities entsprechende Verträge mit Dienstleistern abschließen, auf Regelungen zu achten, die dafür sorgen, dass wissenschaftliche Daten dauerhaft erhalten bleiben.

Daten und Code müssen migrierbar bleiben

Der RfII empfiehlt, insbesondere darauf zu achten, dass Diensteanbieter vertraglich gewährleisten, dass Forschungsdaten bei Vertragsende, im Falle eines Verkaufs des Dienstes bzw. des Dienstangebots oder einer Insolvenz auf eine öffentlich getragene oder geförderte Infrastruktur migrierbar sind und migriert werden dürfen, so dass sie der Wissenschaft weiter zur Verfügung stehen.⁸⁸ Gleiches muss für den Code gelten, in den diese Daten ggf. eingebettet sind.

Lock-In-Effekte ausschließen

Für proprietäre Quellcodes insbesondere von Speicher- und Archivierungssoftware empfiehlt der RfII die Hinterlegung bei Treuhändern, um diese bei Ausfall oder Wechsel des Dienstes zumindest für den Zeitraum der Datenmigration rekonstruieren zu können. Vendor oder Plattform-Lock-Ins müssen im Zuge der Vertragsgestaltung zwischen wissenschaftlichen Akteuren und ihren Partnern ausgeschlossen werden.

5.6 WISSENSCHAFTLICHE QUALITÄTSSICHERUNG GEWÄHRLEISTEN – AUCH AUF SEITEN DES KOOPERATIONSPARTNERS

Empfehlungen 5.6

- Verantwortlichkeiten für und Standards von Qualitätssicherung für eingebrachte Forschungsdaten vertraglich fixieren
- Keine Unterbrechung der wissenschaftlichen Qualitätssicherungsketten zulassen
- Kultur des Explizierens auch von Kooperationspartnern einfordern

⁸⁸ Die Aufwände und technischen sowie personellen Investitionen für eine gelingende Migration wären für die wissenschaftlichen Infrastrukturen hoch. Um hierfür Voraussetzungen schaffen zu können, wären Förderanstrengungen ähnlich dem DFG-geförderten Nationalen Hosting hilfreich. Vgl. dazu S. 64.

Wissenschaftlerinnen und Wissenschaftler sowie wissenschaftliche Einrichtungen, Hochschulen und Konsortien sollten in Kooperationen mit Datendiensten grundsätzlich auf Vereinbarungen bestehen, die die Verantwortlichkeiten für und die Standards der Qualitätssicherung von eingebrachten Forschungsdaten definieren. Die in der Wissenschaft bzw. ihren einzelnen Disziplinen, Feldern und Domänen geltenden Ketten bzw. Zyklen der Qualitätssicherung von Daten dürfen auch dann nicht unterbrochen werden, wenn die Daten an kooperierende Anbieter von Diensten abgegeben werden. Die vom RfII eingeforderte wissenschaftliche Kultur des Explizierens von Daten (etwa durch Protokolle und Metadaten) muss auch auf der Seite des Kooperationspartners ihre Geltung behalten. Der RfII empfiehlt deshalb, auch diese gemäß den wissenschaftlichen Standards fortzusetzenden Qualitätssicherungsschritte vertraglich abzusichern.

Verantwortlichkeiten
festlegen –
Qualitätssicherung
aufrechterhalten

5.7 SELEKTIONS- UND BEWERTUNGSKRITERIEN FÜR FORSCHUNGSDATEN ANHAND WISSENSCHAFTLICHER VORGABEN ENTWICKELN

Empfehlungen 5.7

- Bewertungsaufgaben, die das Priorisieren von Daten nach außerwissenschaftlichen Kriterien begünstigen, nicht an Dienste delegieren
- Mechanismen zur wissenschaftsimmanenten Bewertung der „Reputation“ von Daten und Datensätzen aufbauen
- Metriken zur Datenbewertung behutsam einsetzen

Der RfII empfiehlt Forscherinnen und Forschern, wissenschaftlichen Einrichtungen und Hochschulen sowie Konsortien bei der Auswahl und Abgabe der Daten an kooperierende Dienstleister oder automatisierte Dienste darauf zu achten, dass keine Selektions- oder Bewertungsaufgaben übertragen bzw. quantitative Bewertungsschemata angewendet werden, die nahelegen, dass das Priorisieren bzw. „Ranken“ von Daten nach außerwissenschaftlichen Kriterien Aussagen über die Qualität der Daten erlaubt. Die wissenschaftlichen Akteure tragen dazu bei, dass Mechanismen (z.B. Annotationsformen) aufgebaut werden, die die „Reputation“ von Daten und Datensätzen auf der Grundlage von wissenschaftsimmanenten Bewertungen unterstützen. Grundsätzlich mahnt der RfII, mit automatisierten Findwerkzeugen aus nichtwissenschaftlichen Bereichen sowie Metriken zur Datenbewertung vorsichtig umzugehen: heute „nutzlos“ erscheinende oder einfach temporär nicht genutzte Daten können die Forschung von morgen entscheidend voranbringen oder als wichtige Referenz für künftige Untersuchungen dienen.

Mit Bewertungen
behutsam umgehen
– wissenschaftliche
Kriterien ansetzen

5.8 FACHLICH-INHALTLICHE AUFBEREITUNG DER DATEN FÜR AUßERWISSENSCHAFTLICHE KONTEXTE LEISTEN

Empfehlungen 5.8

- Aussagefähige Forschungsdaten vor der Weitergabe an verwertende bzw. vermarktende Partner redaktionell aufbereiten
- Zusätzlichen Ressourcenbedarf der Wissenschaftsakteure für eine zielgruppenorientierte Datenaufbereitung durch Wissenschaftsakteure decken
- Anerkennung von Datenprodukten als eigenständige wissenschaftliche Leistungen fördern

Datenaufbereitung an fachliche Expertise binden

Aus Sicht des RfII bedarf die gezielte Weitergabe von Daten an verwertende bzw. vermarktende Partner oder an die Öffentlichkeit seitens der Wissenschaft einer fachlich-inhaltlichen Aufbereitung der ursprünglichen Daten (Selektion, Gruppierung, Annotation, Kommentierung) für die neuen, außerwissenschaftlichen Zwecke. Dieser Vorgang wird hier in Anlehnung an publizistische Informationsaufbereitung *Redaktion* (von Daten) genannt. Unterstrichen wird damit die Bedeutung der Rückbindung von Datenaufbereitung an (fach)wissenschaftliche Expertise.

Zielgruppenorientierte Aufbereitung bedarf eines gesondert zu deckenden Ressourceneinsatzes

Eine zielgruppenorientierte Aufbereitung von Forschungsdaten für außerwissenschaftlichen Gebrauch bedarf eines besonderen Aufwandes, der nicht mit der innerwissenschaftlichen Qualitätssicherung gleichzusetzen ist.⁸⁹ Sofern solche Redaktionsprozesse für die Bereitstellung von Forschungsdaten gesetzlich vorgeschrieben werden bzw. vorgeschrieben sind – z.B. eine Bereitstellungspflicht in unterschiedlichen Sprachen oder interessentengerechten Datenformaten, die eine unmittelbare Verwertung oder Vermarktung durch Dritte ermöglichen – wird absehbar ein höherer Einsatz an Zeit, Personal und technischer Infrastruktur in den Einrichtungen entstehen. Entsprechend wird der hierfür aufzuwendende Ressourcenbedarf wachsen und ist durch hierauf bezogene Mittelzuweisungen an Hochschulen und Forschungsorganisationen sowie bei Allokationsentscheidungen innerhalb der Einrichtungen zu berücksichtigen. Entsprechende Leistungen und für diese Form der Leistungserbringung qualifiziertes Personal können in der Regel nicht aus den bisherigen institutionellen Etats von Hochschulen und außeruniversitären Forschungseinrichtungen finanziert werden. Um nachträglich entstehenden Kostenaufwand gering zu halten, empfiehlt der RfII den Forschungsakteuren, Möglichkeiten einer späteren Datennachfrage bzw. Datenverwendung auch durch zivilgesellschaftliche und kommerzielle Interessenten bereits in der Projektplanung mit zu bedenken und frühzeitig entsprechende Vorkehrungen zu treffen. Die Finanzierung

⁸⁹ Vgl. dazu RfII (2019) – Herausforderung Datenqualität, 75 f.

entsprechender Bedarfe müsste dann auch bereits in der Projektförderung berücksichtigt werden.

Der wissenschaftlichen Redaktion wie auch schon der Erhebung vorgelagert sind Verständigungsprozesse in den Fachgemeinschaften, Forschungsfeldern und Domänen, welche Daten überhaupt für außerwissenschaftliche Verwendung und Vermarktung geteilt werden sollten. Als Foren für entsprechende Aushandlungsprozesse sieht der RfII auf nationaler Ebene zum Beispiel die Konsortien der NFDI als geeignet an.

Durch die Redaktion entstehen in der Wissenschaft „Datenprodukte“, die für außerwissenschaftliche Interessenten von unmittelbarem Verwertungs- bzw. Vermarktungsinteresse sind. Diese Datenprodukte stellen nach Auffassung des RfII zudem eigenständige wissenschaftliche Leistungen dar, die im Rahmen von Verfahren zur Reputationszuweisung und Leistungsbewertung (z.B. Berufungen, Preisverleihungen, Evaluationen und Akkreditierungen) sowie bei indikatoren-gestützten Mittelzuweisungen zu berücksichtigen sind.⁹⁰

Datenprodukte
als eigenständige
wissenschaftliche
Leistungen
anerkennen

EMPFEHLUNGEN ZUR (SELBST-)ORGANISATION VON HANDLUNGSFÄHIGKEIT AUF SEITEN DER WISSENSCHAFT

5.9 BERATUNG UND VERMITTLUNG FÜR FACHGEMEINSCHAFTEN ORGANISIEREN – EINE CHANCE FÜR DIE NFDI UND IHRE KONSORTIEN

Empfehlungen 5.9

- Weiterentwicklung der NFDI und ihrer Konsortien zu Beratern und Mittlern der Communities bei der Beauftragung externer Datendienste forcieren
- Kompetenzaufbau in Rechtsfragen und Vertragsgestaltung zur Ausgestaltung symmetrischer Partnerschaften mit Diensteanbietern in der Wissenschaft gewährleisten
- Die Machtposition von Communities, Hochschulen und Wissenschaftseinrichtungen in Verhandlungen mit (kommerziellen) Diensteanbietern stärken

Mit der Nationalen Forschungsdateninfrastruktur (NFDI) haben Bund und Länder eine ortsverteilte Struktur etabliert, in der drängende Fragen zu digitalen Forschungsdaten und ihren Infrastrukturen seit 2020 gebündelt werden. Kompetenz wird dabei auch über akademische Fächergrenzen hinweg zusammengebracht.

Ansprechpartnerin
und Mittlerin
im Umgang mit
Diensteanbietern: Eine neue
Rolle für die NFDI?

⁹⁰ Vgl. dazu Empfehlung 4.4.1 RfII (2019) – Herausforderung Datenqualität, 88 ff.

Aus Perspektive des RfII ist die NFDI mit ihren Konsortien deshalb für Einzelwissenschaftlerinnen und-wissenschaftler sowie für Verbundprojekte und andere Forschungszusammenschlüsse auch für die Suche nach geeigneten externen Dienstangeboten ein geeigneter Ansprechpartner. Sofern im Rahmen der NFDI selbst entsprechende Dienste (noch) nicht zugänglich sind bzw. durch die NFDI absehbar gar nicht aufgesetzt oder initiiert werden, sollten die NFDI-Konsortien (bei fachspezifischen Diensten) oder die gesamte NFDI (bei generischen Diensten) die Rolle der Beratung und Vermittlung übernehmen.

Rechtliche
Kompetenzen
aufbauen – Verträge
auf Augenhöhe
verhandeln

In diesem Zusammenhang betont der RfII, dass neben Kompetenzen in der Beratung, die die fachliche Eignung eines Dienstes betreffen, auch Kompetenzen hinsichtlich der rechtlichen Beratung zur Ausgestaltung möglichst symmetrischer Partnerschaften – „auf Augenhöhe“ – bereitgestellt werden müssen. Guten Vertragskonstruktionen kommt für eine Souveränität (digitaler) Wissenschaft eine Schlüsselrolle zu. Die NFDI kann auch das Instrument sein, sich zum Zweck von Vertragsabschlüssen zusammenzutun. Diese Möglichkeit können nach Bedarf einzelne oder mehrere Konsortien nutzen aber auch die NFDI in ihrer Gesamtheit. Ziel ist es hierbei, Forschungs-Communities, Hochschulen und Wissenschaftseinrichtungen zu ertüchtigen, aus einer eigenen starken Machtposition heraus mit (kommerziellen) Diensteanbietern verhandeln zu können. Flankiert von den notwendigen staatlichen Rahmenseetzungen (Empfehlungen 5.1 bis 5.3) würde ein solcher aus der Wissenschaft selbst kommender Beitrag zur Ermöglichung von Verhandlungsmacht wesentlich dazu beitragen, die Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen und einer spezifisch wissenschaftlichen Datenkultur voran zu bringen. Hierbei versteht sich, dass die Vermeidung von (ggf. auch neuen) Oligopolen und Quasi-Oligopolen gerade bei Vertragslösungen auf nationaler oder EU-Ebene unverändert Priorität haben muss.

5.10 ESSENTIELLE DATENBESTÄNDE DER WISSENSCHAFTLICHEN FACHGEMEINSCHAFTEN UND DOMÄNEN DEFINIEREN UND SICHERN

Empfehlungen 5.10

- Kernbestände von Forschungsdaten variabel und zukunfts offen festlegen und sichern
- Die Entwicklung wissenschaftseigener Dienste für den „kritischen“ Datenbereich von Communities und Fächern erwägen
- Rechtsformen zur wissenschaftskonformen Ausgestaltung von Diensten an den fachlichen oder domänenbezogenen Zwecken einer (gebührenfreien) Bereitstellung und Nutzung orientieren

Im Handlungsfeld der NFDI-seitigen Beratung und Vermittlung für die wissenschaftlichen Fachgemeinschaften sieht der RfII auch die Bestimmung und Sicherung essentieller wissenschaftlicher Datenbestände. Unter essentiellen Datenbeständen sollen hier Daten verstanden werden, die unter fachlichen Gesichtspunkten bezüglich Aufbewahrung und Speicherung „Ewigkeitscharakter“ haben; also Daten, die im Zentrum der Methodenentwicklung und Theoriebildung eines Faches, Forschungsfeldes oder einer Domäne stehen. Hier ist insbesondere zu berücksichtigen, dass Konsensbildung über einen solchen „Kernbestand“ keine leichte Aufgabe ist. Im Zuge des wissenschaftlichen Fortschritts ist auch in dieser Hinsicht Variabilität geboten. Die Kanonisierung von Beständen, die Bildung von Referenzkorpora und ihre Zugänglichkeit können Forschungslinien von Disziplinen über Jahrzehnte prägen – dabei gleichzeitig aber andere Forschungslinien ungewollt abschneiden. Insofern ist auch die Frage nicht trivial, welcher Dienstleister mit der Aufbewahrung und Speicherung entsprechender digitaler Datenbestände oder – weitergehender – mit ihrer Kuratierung, Annotierung und transdisziplinären Bereitstellung (in anschlussfähigen Formaten) beauftragt werden soll.

Konsensbildung über
Kerndatenbestände
organisieren

Der RfII empfiehlt frühzeitig, etwa im Rahmen der NFDI, aber auch in anderen Fachforen und den DFG-Fachkollegien den diesbezüglichen Optionenraum auszu-leuchten. Neben einer Beauftragung wissenschaftsexterner Anbieter mit (auch) kommerziellen Interessen sollten wissenschaftseigene Lösungen erwogen werden, die je nach Bedarf und Eignung, in einem breiten gesellschaftsrechtlichen Spektrum organisiert werden können – vom eingetragenen Verein, über gemeinnützige GmbHs bis hin möglicherweise zu Datengenossenschaften als neuer Rechtsform.

Aufbau
wissenschaftseigener
Dienste erwägen –
geeignete
Rechtsformen
identifizieren

6 LITERATURVERZEICHNIS

- Anderson, Warwick et al. (2017): Towards Coordinated International Support of Core Data Resources for the Life Sciences, 7 S., DOI: 10.1101/110825.
- Arbeitsgruppe Virtuelle Forschungsumgebungen in der Allianz der Deutschen Wissenschaftsorganisationen (2011): Definition Virtuelle Forschungsumgebung, DOI: 10.2312/ALLIANZOA.028.
- Assante, Massimiliano et al. (2016): Are Scientific Data Repositories Coping with Research Data Publishing?, in: Data Science Journal 15, Nr. 6, S. 1-24, DOI: 10.5334/dsj-2016-006.
- AWBI- Ausschuss für Wissenschaftliche Bibliotheken und Informationssysteme (2021): Datentracking in der Wissenschaft: Aggregation und Verwendung bzw. Verkauf von Nutzungsdaten durch Wissenschaftsverlage. Ein Informationspapier des Ausschusses für Wissenschaftliche Bibliotheken und Informationssysteme der Deutschen Forschungsgemeinschaft 20. Mai 2021. Hg. v. DFG- Deutsche Forschungsgemeinschaft, Bonn, 13 S., online verfügbar unter: https://www.dfg.de/download/pdf/foerderung/programme/lis/datentracking_papier_de.pdf, zuletzt geprüft am: 20.06.2021.
- Bernard, Lars/Mäs, Stephan (2018): e-science in den Geo- und Umweltwissenschaften. Ansätze und Anforderungen, in: Informatik Spektrum 94, Nr. 1, S. 112, DOI: 10.1007/s00287-018-01138-x.
- Brembs, Björn/Förstner, Konrad et. al. (2020): Überwachte Wissenschaftler. Auf einmal Laborratte, in: Frankfurter Allgemeine Zeitung, online verfügbar unter: <https://www.faz.net/aktuell/karriere-hochschule/hoersaal/auf-einmal-laborratte-wie-grossverlage-wissenschaftler-ueberwachen-17078766.html>, zuletzt geprüft am: 02.06.2021.
- Bundesregierung (2021): Entwurf eines Gesetzes zur Änderung des E-Government-Gesetzes und zur Einführung des Gesetzes für die Nutzung von Daten des öffentlichen Sektors, 52 S., online verfügbar unter: <https://www.bmwi.de/Redaktion/DE/Downloads/> Gesetz/gesetzentwurf-egov-open-data-gesetz-und-datennutzungsgesetz.pdf?__blob=publicationFile&v=4, zuletzt geprüft am: 11.02.2021.
- DFG- Deutsche Forschungsgemeinschaft (2020): Digitaler Wandel in den Wissenschaften. Impulspapier. Unter Mitarbeit von Matthias Katerbow et al., 16 S., DOI: 10.5281/zenodo.4191345.
- Dreier, Thomas et al. (Hg.) (2016): Informationen der öffentlichen Hand: Zugang und Nutzung, Baden-Baden: Nomos, 596 S.
- EC- Europäische Kommission (2020): Eine europäische Datenstrategie. Mitteilung der Kommission an das Europäische Parlament, den Rat, den Europäischen Wirtschafts- und Sozialausschuss und den Ausschuss der Regionen, Brüssel, 39 S., online verfügbar unter: https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_de.pdf, zuletzt geprüft am: 03.04.2021.
- EC- European Commission (2019): Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on Open Data and the Reuse of Public Sector Information (Recast). Directive (EU) 2019/1024, vom: 20.06.2019. Official Journal of the European Union, 28 S., online verfügbar unter: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=uriserv:OJ.L_.2019.172.01.0056.01.ENG, zuletzt geprüft am: 04.07.2021.
- EFI- Expertenkommission Forschung und Innovation (2016): Gutachten 2016. Gutachten zu Forschung, Innovation und technologischer Leistungsfähigkeit Deutschlands, Berlin, 172 S., online verfügbar unter: http://www.e-fi.de/fileadmin/Gutachten_2016/EFI_Gutachten_2016.pdf, zuletzt geprüft am: 18.12.2018.
- ELIXIR (2017): Public Data Resources as a Business Model for SMEs. The Role of Public Bioinformatics Infrastructure in Supporting Innovation in the Life Sciences, Hinxtion, 15 S., online verfügbar unter: <https://elixir-europe.org/open-data-SMEs-report>, zuletzt geprüft am: 04.06.2021.

Frank, Gwen (2019): A look into Figshare with Mark Hahnel, CEO. Insights into the Economy of Open Scholarship. Hg. v. Knowledge Exchange, 12 S., online verfügbar unter: http://repository.jisc.ac.uk/7296/4/KE_Inisghts_A_look_into_Figshare_June_2019.pdf, zuletzt geprüft am: 17.08.2020.

Heyde, Markus von der (2019): Open Research Data. Landscape and Cost Analysis of Data Repositories Currently Used by the Swiss Research Community, and Requirements for the Future, Weimar, 67 S., online verfügbar unter: DOI: 10.5281/zenodo.2643430, zuletzt geprüft am: 04.06.2020.

Hohls, Rüdiger/Prinz, Claudia/Schlottheuber, Eva (Hg.) (2016): Historische Grundwissenschaften und die digitale Herausforderung (Historisches Forum, 18), H-Soz-Kult, Berlin, 57 S., DOI: 10.18452/18771.

KE- Knowledge Exchange (2014): Sowing the Seed. Incentives and Motivations for Sharing Research Data, a Researcher's Perspective, Kopenhagen, 48 S., online verfügbar unter: http://repository.jisc.ac.uk/5662/1/KE_report-incentives-for-sharing-researchdata.pdf, zuletzt geprüft am: 04.06.2021.

Klaffki, Lisa/Schmunk, Stefan/Stäcker, Thomas (2018): Stand der Kulturgutdigitalisierung in Deutschland. Eine Analyse und Handlungsvorschläge des DARIAH-DE Stakeholdergremiums „Wissenschaftliche Sammlungen“ (DARIAH-DE Working Papers, 26), 39 S., URN: urn:nbn:de:gbv:7-dariah-2018-1-3.

Konrad, Uwe et al. (2020): Digitale Dienste für die Wissenschaft- wohin geht die Reise? Positionspapier. Unter Mitarbeit von Mathias Bornschein et al., 36 S., DOI: 10.5281/ZENODO.4301924.

KVI- Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (Hg.) (2001): Wege zu einer besseren informationellen Infrastruktur. Gutachten der vom Bundesministerium für Bildung und Forschung eingesetzten Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik. Inhaltsverzeichnis, Baden-Baden: Nomos, 297 S.

Lauber-Rönsberg, Anne/Krahn, Philipp/Baumann, Paul (2018): Gutachten zu den rechtlichen Rahmenbedingungen des Forschungsdatenmanagements, 23 S., online verfügbar unter: https://tu-dresden.de/gsw/jura/igetem/jfbimd13/ressourcen/dateien/publikationen/DataJus_Zusammenfassung_Gutachten_12-07-18.pdf?lang=de, zuletzt geprüft am: 02.06.2021.

Neylon, Cameron et al.: Open Scholarship and the Need for Collective Action, 102 S., DOI: 10.5281/ZENODO.3454687.

OECD- Organisation for Economic Cooperation and Development (2017): Business Models for Sustainable Research Data Repositories (OECD Science, Technology and Innovation Policy Papers, 47), 80 S., DOI: 10.1787/302b12bb-en.

Open Working (2020): Why figshare? Choosing a new technical infrastructure for 4TU.ResearchData, online verfügbar unter: <https://openworking.wordpress.com/2020/08/18/why-figshare-choosing-a-new-technical-infrastructure-for-4tu-researchdata/>, zuletzt geprüft am: 02.06.2021.

Pappenberger, Karlheinz (2016): bwFDM-Communities – Wissenschaftliches Datenmanagement an den Universitäten Baden-Württembergs, in: Bibliothek Forschung und Praxis 40, Nr. 1, S. 21-25, DOI: 10.1515/bfp-2016-0017.

RfII- German Council for Scientific Information Infrastructures (2020): Comment on the UNESCO Open Science Initiative, 4 S., URN: urn:nbn:de:101:1-2020052620.

RfII- Rat für Informationsinfrastrukturen (2016): Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland, Göttingen, 160 S., URN: urn:nbn:de:101:1-201606229098.

RfII- Rat für Informationsinfrastrukturen (2019): Herausforderung Datenqualität – Empfehlungen zur Zukunftsfähigkeit von Forschung im digitalen Wandel, 2. Aufl., Göttingen, 172 S., URN: urn:nbn:de:101:1-2020043009455657544963.

Rfll- Rat für Informationsinfrastrukturen (2019): Stellungnahme des Rfll zu aktuellen Entwicklungen rund um Open Data und Open Access, Göttingen, 8 S., URN: urn:nbn:de:101:1-2019051612340151989161.

Rfll- Rat für Informationsinfrastrukturen (2020): Datentreuhandstellen gestalten – Zu Erfahrungen der Wissenschaft. Stellungnahme, Göttingen, 8 S., URN: urn:nbn:de:101:1-2020043009112405568503.

Rfll- Rat für Informationsinfrastrukturen (2020): Datendienste nachhaltig gestalten. Ein Diskussionsimpuls zur Weiterentwicklung von Forschungsdateninfrastrukturen, Göttingen, 6 S., URN: urn:nbn:de:101:1-2020071501211536559967.

Rfll- Rat für Informationsinfrastrukturen (2021): Stellungnahme zum Vorschlag eines Data Governance Acts (DGA) durch die EU-Kommission, Göttingen, 6 S., URN: urn:nbn:de:101:1-2020052654.

RSC- Royal Society of Chemistry (2009): RSC acquires ChemSpider, online verfügbar unter: <https://www.rsc.org/news-events/articles/2009/05-may/chem-spider/>, zuletzt geprüft am: 02.06.2021.

Soltau, Kerstin/Razum, Matthias/Strecker, Dorothea (2019): Management digitaler Forschungsdaten im akademischen Umfeld – Lessons learned aus der Einführung von RADAR, in: Mittermaier, Bernhard (Hg.): Forschungsdaten. Sammeln, sichern, strukturieren (Schriften des Forschungszentrums Jülich. Reihe Bibliothek, 23), S. 119-135, online verfügbar unter: https://user.fz-juelich.de/record/863157/files/3.1_Soltau_RADAR.pdf, zuletzt geprüft am: 04.06.2021.

Teal, Tracy et al. (2020): Sustainable, Open Source Alternatives Exist. Hg. v. Dryad news and views, online verfügbar unter: <https://blog.datadryad.org/2020/08/26/sustainable-infrastructure-exists/>, zuletzt geprüft am: 04.06.2021.

United States District Court Southern District of New York (2016): Jury Trial Demanded. Highsmith v. Getty Images 16 Civ. 5924, 38 S., online verfügbar unter: <https://assets.documentcloud.org/documents/2999>

596/Gov-UScourts-Nysd-460787-1-0.pdf, zuletzt geprüft am: 10.05.2021.

Verbraucherzentrale Bundesverband e.V. (2021): Vertrauen stärken durch verbraucherfreundliche Daten-Governance. Stellungnahme des Verbraucherzentrale Bundesverbands zum Vorschlag der EU-Kommission für eine Verordnung über europäische Daten-Governance, Berlin, online verfügbar unter: https://www.vzbv.de/sites/default/files/downloads/2021/01/13/21-01-12_vzbv-stellungnahme_data-governance-act.pdf.

Whyte, Angus/Pryor, Graham (2011): Open Science in Practice: Researcher Perspectives and Participation, in: IJDC- International Journal of Digital Curation 6, Nr. 1, S. 199-213, DOI: 10.2218/ijdc.v6i1.182.

WR- Wissenschaftsrat (2017): Empfehlungen zur Weiterentwicklung der Deutschen Zentren der Gesundheitsforschung, Köln, 153 S., online verfügbar unter: <https://www.wissenschaftsrat.de/download/archiv/6413-17.pdf>, zuletzt geprüft am: 02.06.2021.

WR- Wissenschaftsrat (2020): Zum Wandel in den Wissenschaften durch datenintensive Forschung. Positionspapier, 72 S., online verfügbar unter: https://www.wissenschaftsrat.de/download/2020/8667-20.pdf?__blob=publicationFile&v=4, zuletzt geprüft am: 01.06.2021.

WR- Wissenschaftsrat (2021): Impulse aus der COVID-19-Krise für die Weiterentwicklung des Wissenschafts-systems in Deutschland. Positionspapier, Köln, 73 S., online verfügbar unter: https://www.wissenschaftsrat.de/download/2021/8834-21.pdf?__blob=publication-File&v=15, zuletzt geprüft am: 28.05.2021.

ANHANG

A. ALPHABETISCHES VERZEICHNIS DER FALLBEISPIELE

1. *Arachne, iDAI.objects arachne*

Domänenspezifische Bild- und Objektdatenbank, die nach anfänglich institutioneller Finanzierung (Beginn 1995) und einer DFG-Förderung (2009-2013) als Teil des iDAI.welt durch das Deutsche Archäologische Institut (DAI) ausfinanziert ist. Betreiber sind das DAI und das CoDArchLab der Universität zu Köln. Bei der technischen Entwicklung kooperiert DAI mit dem Lehrstuhl für Historisch-Kulturwissenschaftliche Informationsverarbeitung an der Universität zu Köln. Der Dienst bietet eine Datenbank, Werkzeuge für Recherche, Erschließung der analogen und digitalen Bestände, 3D-Visualisierung sowie Thesauri. Der Basisdienst ist kostenfrei. Durch eine Registrierung erhalten die Nutzenden die digitalen Objekte in höherer Auflösung. Für die Erstellung verwertbarer Abzüge werden Bearbeitungsgebühren erhoben. Die Inhalte sind ggf. urheberrechtlich geschützt. Das Bildmaterial wird ausschließlich für wissenschaftliche Zwecke zur Verfügung gestellt.

arachne.uni-koeln.de/drupal

2. *Bavarikon – Internetportal Kultur und Wissensschätze Bayerns*

Das Internetportal stellt „Kunst-, Kultur- und Wissensschätze“ aus bayerischen Archiven, Bibliotheken und Museen sowie Institutionen der Landesverwaltung, der Denkmalpflege und aus der Wissenschaft bereit. Die Digitalisate wurden teilweise eigens dafür produziert. Fördermittel stammen aus dem Programm „Bayern Digital“. Über die Finanzierung und grundsätzliche Ausrichtung entscheiden die Staatsministerien für Wissenschaft und Kunst bzw. für Digitales; ein Rat, bestehend aus 13 Mitgliedern bayerischer Institutionen, entscheidet über die Aufnahme neuer Inhalte und berät die Leitungsebene zur Digitalisierungsstrategie. Technischer Dienstleister ist das Leibniz-Rechenzentrum der Bayerischen

Akademie der Wissenschaften. Da die Rechte an den Objekten bei den beitragenden Einrichtungen liegen, ist die Nutzung der Inhalte nicht einheitlich geregelt und muss individuell verhandelt werden.

bavarikon.de

3. *Beck-online*

Der Verlag C.H. Beck hat mit seiner 2011 errichteten juristischen online Datenbank eine marktführende Position erlangt, indem er seine klassischen Printmedien für juristische Wissenschaften für die Onlinenutzung optimiert und diese durch ein breites Angebot von Content-Partnern ergänzt: es bestehen Kooperationen mit Gerichten, Verlagen, Kanzleien, Verbänden, Vereinen, Instituten, öffentlichen Einrichtungen. Dabei werden durch Einfügen von Leitsätzen, Normenketten, Schlagworten sowie Angaben zum Verfahrensgang freie Daten aus der öffentlichen Hand kommerzialisiert. Die Finanzierung erfolgt über eine monatliche Pauschale, die sich nach den lizenzierten Fachmodulen richtet. Inhalte, die nicht vom Modul umfasst sind, können einzeln erworben werden. Das reine Durchsuchen der Datenbank ist nach einer Registrierung jedoch kostenlos. Mit den Bundesländern wurden Vereinbarungen über die Nutzung in der Verwaltung und in der Justiz getroffen. Bibliotheken können ein Hochschulmodul lizenzieren.

beck-online.beck.de

4. *Chemical Abstracts Service (CAS)*

CAS ist seit 1907 Herausgeber der Chemical Abstracts (CA), eines Publikationsorgans der American Chemical Society. Es werden Datenbanken erstellt, eine CAS Registry Number vergeben sowie weltweit Abstracts sämtlicher chemie-relevanter Veröffentlichungen gesammelt. Dadurch ist CAS für Forschung, Wirtschaft

und Behörden ein zentrales Referenzkorpus. Die Aufnahme der Daten geschieht nach wissenschaftlichen Gesichtspunkten. Das Angebot umfasst neben einer Datenbank die Suche, Beschreibung und Organisation von Experimenten und Analysetechniken, Daten zu aktuellen Forschungsprojekten, Patenten, historischen chemischen Zeitschriften u. Ä. Für die Industrie wird zusätzlich eine Unterstützung bei kritischen Geschäftsentscheidungen angeboten. Dienstleistungen und Produkte sind überwiegend kommerziell, auch das Programm SciFinder zum Durchsuchen der Datenbanken wird bepreist. Eine in den Suchmöglichkeiten eingeschränkte Version, die häufig an Universitäten eingesetzt wird, ist gleichfalls lizenzpflichtig. Das FIZ Karlsruhe ist der Servicepartner für alle europäischen Kunden.

cas.org

5. CERA – Langzeitarchiv des Deutschen Klimarechenzentrums

Das Archiv des Deutschen Klimarechenzentrums (DKRZ) ist auf Langzeitbewahrung ausgelegt. Aufgenommen werden Daten aus HPC Projekten des DKRZ sowie externe Ressourcen. Forschende schließen dazu einen Vertrag mit dem Repositorium. Die archivierten Datensätze können auch publiziert werden (CC-Lizenzen). Verschiedene Harvester wie Google Dataset Search und World Data System werden vom Langzeitarchiv mit Metadaten beliefert. Für die laufende Forschung (Projekt- und Analysephase) gibt es mit der Earth System Grid Federation (ESGF) ein gesondertes Angebot, das den Plattformdiensten ähnelt, die auch im Bereich der Satellitendaten entstehen (Daten in Kombination mit Applikationen und Cloud-basierten Speicher- und Rechenkapazitäten). Die wissenschaftliche Nachnutzung der Daten ist kostenlos, eine wirtschaftliche Verwertung ist nach Vereinbarung möglich. Eine Quellenangabe

ist bei der Nutzung erforderlich. Der Dienst wird vom NFDI4Earth-Konsortium genutzt.

cera-www.dkrz.de

6. ChemSpider

Entwickelt als Privatprojekt, wird die kostenfreie, 103 Mio. chemische Strukturen und 276 Datenquellen enthaltende Datenbank seit 2009 von der Royal Society of Chemistry gehostet. Die Kuratierung erfolgt sowohl durch einige wenige hauptamtliche Kuratorinnen, als auch durch die Fachcommunity (Korrekturen, Annotationen und Ergänzungen). Die heterogenen Daten sind frei zugänglich, über ein Discovery System auffindbar und mit zusätzlichen Informationen und Eigenschaften ergänzt sowie mit den ursprünglichen Datenquellen verlinkt. Die Finanzierung erfolgt über Anzeigen, Spenden und Kooperationen mit der Wirtschaft. Das NFDI4Chem-Konsortium sieht den Dienst als eine Möglichkeit für die Vernetzung etablierter Repositorien.

chemspider.com

7. Copernicus – Europäische Sentinel-Daten

Das COPERNICUS Erdbeobachtungsprogramm der europäischen Mitgliedsstaaten (European Earth Observation and Monitoring Programme) beinhaltet die Verfügbarmachung der Daten durch die European Space Agency ESA – explizit auch mit dem Ziel der gewerblichen Nutzung bzw. Vermarktung. Es handelt sich um Big Data, die bei den Nutzenden eine entsprechende IT-Infrastruktur bzw. Rechenkapazität erfordert. Für den Download ist eine Registrierung nötig. Die Sentinel Daten können via Copernicus Open Access Hub oder Copernicus Online Data Access verwendet werden. Zudem steht seit Juni 2018 ein Zugang über DIAS Data and Information Access Service zur Verfügung. Die Vermarktung der Daten umfasst daher oftmals auch

technologische Angebote wie Software-as-a-Service und Infrastructure-as-a-Service. Der Aufbau dieser Vermarktungsplattformen für Daten und Dienste wird langfristig mit europäischen Fördermitteln unterstützt und auf der nationalen Ebene durch eine „Nationale Strategie“ verstetigt.

copernicus.eu

8. Deutsche Digitale Bibliothek (DDB)

Auf Basis eines öffentlichen Auftrags (Beschluss der Bundesregierung) aggregierte (Meta)Daten aus der Wissenschaft sowie von Kultureinrichtungen und Behörden können sektorenübergreifend genutzt werden. Es handelt sich um ein Zugangsportal zu Daten von den im Kompetenznetzwerk kooperierenden Kultur- und Wissenschaftseinrichtungen: Als potentielle Datenerzeuger haben sich über 2000 Institutionen registriert. Gewährleistet wird ein offener, kostenloser Zugang zu Metadaten und Erschließungsinformationen sowie zu Auszügen oder Kleinformaten, wie z.B. Inhaltsverzeichnissen, Vorschaubildern, Audio- und Videoausschnitten. Es gelten Nutzungsbedingungen der gebenden Einrichtung. Ein besonderer Adressat ist hier die Zivilgesellschaft. Der technische Betrieb erfolgt durch das FIZ Karlsruhe. Zur Governance gehören Mitgliederversammlung, Vorstand und Kuratorium. Der Dienst wird von Bund und Ländern auf der Grundlage eines Verwaltungs- und Finanzierungsabkommens finanziert.

deutsche-digitale-bibliothek.de

9. Deutsches Referenzkorpus (DeReKo)

DeReKo ist ein vom Leibniz-Institut für Deutsche Sprache betriebener, über Drittmittelprojekte und aus institutionellen Mitteln finanzierter Datendienst für empirische linguistische Forschung. Die inhaltlich kuratierte Sammlung gegenwärtiger

deutschsprachiger Korpora steht für wissenschaftliche Nutzung kostenfrei bereit. Die nicht urheberrechtlich geschützten Teile der Korpora sind öffentlich zugänglich. Für wissenschaftliche Nutzung gibt es mit den gebenden Einrichtungen (beispielsweise Verlagen) Lizenzvereinbarungen für weitere Korpora. Eine Registrierung ermöglicht Zugang zu solchen Beständen. Neben der Datenbank wird eine Korpusanalyseplattform angeboten. Die Bestandserweiterung erfolgt mit Blick auf inhaltliche Erweiterung, Umfang und Qualität, aber auch aufgrund der Beschaffungskosten der notwendigen Nutzungsrechte sowie Konvertierungsaufwands. Kommerzielle Nutzung ist ausgeschlossen. Der Dienst wird im Text+ Konsortium eine Rolle spielen.

ids-mannheim.de/digspra/kl/projekte/korpora/

10. Deutsche Fotothek

Der Dienst wird im Rahmen eines Sammelauftrags betrieben, den die SLUB Dresden schon 1924 erhielt. Er stellt über eine Plattform (aufgebaut mit Mitteln des Europäischen Fonds für regionale Entwicklung) zwei Millionen Bilder aus 90 Institutionen bereit. Die Nutzung wird durch Verwertungs- und Urheberrechte (auch für Metadaten) beschränkt. Metadaten sowie die nicht geschützten Bestände sind in Vorschauqualität öffentlich zugänglich, aber ohne Zustimmung nicht nachnutzbar. Die Verwertung ist kostenpflichtig und bedarf vorab der ausdrücklichen Zustimmung. Für private und wissenschaftliche Zwecke, inklusive Printprodukte in kleinen Auflagen und Websites besteht keine Kosten-, aber Belegpflicht. Die institutionelle Finanzierung wird durch Drittmittelprojekte zur Erschließung der einzelnen Bestände ergänzt. Ausdifferenzierte Bereitstellungs- bzw. Erstellungsgebühren für unterschiedliche Nutzungsszenarien werden bei eigenen Beständen erhoben. Nutzungsbedingungen können, je nach datenerzeugender Einrichtung, variieren und müssen jeweils

einzelnen erfragt werden. Die Deutsche Fotothek stellte 2009 für Wikimedia Commons 250.000 Fotos unter CC-SA Lizenz online. Die Deutsche Fotothek ist in das Portfolio des NFDI4Culture-Konsortiums eingebunden.

deutschefotothek.de

11. Deutsches Textarchiv (DTA)

Das Deutsche Textarchiv (DTA) der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW) bietet einen Referenzkorpus mit disziplin- und gattungsübergreifenden, nicht mehr urheberrechtlich geschützten Texten deutscher Sprache in unterschiedlichen, auch strukturierten, maschinenlesbaren Formaten an. Die Auswahl der in den Kernkorpus aufzunehmenden Texte wird nach wissenschaftlichen Gesichtspunkten, auf Grundlage einer von Akademiemitgliedern erstellten und kommentierten Bibliographie getroffen. Eine Ergänzung des Kernkorpus durch die Integration von Texten und Textsammlungen aus nicht eigenen Quellen ist möglich, dazu ist neben der Begründung der wissenschaftlichen Relevanz auch eine Anpassung an das DTA-Basisformat notwendig. Die Finanzierung erfolgt durch eine Reihe von Drittmittelprojekten und die Einbindung in die ERIC- sowie voraussichtlich durch das Text+ Konsortium in die NFDI-Infrastruktur. Der Lesezugriff ist öffentlich, ein Gesamtdownload unbeschränkt möglich, Analyse-tools sind eingebunden.

deutschestextarchiv.de

12. Detherm

Die numerische Datenbank für thermophysikalische Stoffdaten für die Auslegung und das Design von chemischen Apparaten, Anlagen und Prozessen beinhaltet 11,48 Millionen Datensätze – auch Datenbanken von Universitäten. Der Dienst wird vom DECHEMA e.V. (Gesellschaft

für Chemische Technik und Biotechnologie, finanziert durch Mitgliedsbeiträge, eine Stiftung und Fördermittel) in Kooperation mit der DDBST GmbH (Ausgründung Universität Dortmund) betrieben. Es werden eigene Daten der Betreiber angeboten. Die Datenbank wird – von Fachpersonal betreut – jährlich aktualisiert. Die Suche in der Datenbank ist für Registrierte kostenlos. Entgelte werden für Datentabellen erhoben und richten sich nach der Anzahl ihrer Zeilen. Für Forschung und Lehre werden Rabatte bis zu 90 Prozent gewährt. Der Betreiber ist Sprecher des NFDI4Cat-Konsortiums.

dechema.de/Medien/Datenbanken/Detherm.html

13. Digitales Wörterbuch der Deutschen Sprache (DWDS)

Das DWDS ist eine Wortdatenbank, die auf dem Wörterbuch der deutschen Gegenwartssprache aufbaut (Referenzkorpora, Zeitungskorpora, Webkorpora, Spezialkorpora). Sie wird von der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW) im Rahmen des Bund-Länder geförderten Akademienprogramms der Union der deutschen Akademien der Wissenschaften entwickelt und betrieben. Die Finanzierung erfolgt als Drittmittelprojekt auch im Rahmen des vom BMBF finanzierten Zentrums für digitale Lexikographie der deutschen Sprache (ZDL). Korpora werden nach wissenschaftlichen Gesichtspunkten in Hinblick auf Qualität und Aktualität akquiriert. Die Nutzung der lizenzfreien Bestände und Analysewerkzeuge ist kostenfrei, automatisierte Abfragen sind nur mit ausdrücklicher Genehmigung gestattet, Urheberrechte Dritter sind zu beachten, Ergebnisse müssen mit einer Quellenangabe versehen werden. Ressourcen, für die eine Registrierung notwendig sind, dürfen nicht zu kommerziellen Zwecken verwendet werden. Automatisierte Abfragen erfordern eine Genehmigung. Das DWDS wird über Einbindung in Drittmittelprojekte, in ERIC- sowie

voraussichtlich in die NFDI-Infrastruktur durch das Text+ Konsortium finanziert. Die Verfügbarkeit der Korpusdaten und Softwaredienste ist über das CLARIN-Servicezentrum der BBAW zunächst bis Dezember 2024 sichergestellt.

dwds.de

14. Daten für Forschung (DaFo)

Das Münchener Digitalisierungszentrum (MDZ) bietet seit 2016 mit DaFo einen kostenlosen Downloadservice für hochauflösende Seitenabbilder von Druckwerken (hauptsächlich) der Bayerischen Staatsbibliothek, teilweise mit dazugehörigen OCR-Texten. Für die Forschung besteht bei Veröffentlichungen Herkunftsnachweis und Belegexemplarpflicht. Kommerzielle Nutzung ist genehmigungs- und gebührenpflichtig. Für Objekte gelten individuelle Lizenzen.

dafo.digitale-sammlungen.de/dafomain.php

15. Dryad Digital Repository

Das generische Repositorium Dryad wurde im Aufbau durch die National Science Foundation (2007–2019), teilweise über Horizon2020 (2015–2018) sowie JISC (2010–2011) gefördert. Dryad entstand als ein Repositorium für Daten aus Lebenswissenschaften, die mit begutachteten Veröffentlichungen assoziiert sind (auch Daten zu PhD-Thesen, seltener zu Monografien). Es wurde nach Überführung in einen gemeinnützigen Verein nach US-Recht für Daten aus weiteren Fachbereichen geöffnet. Dryad basiert auf einem weltweit offenen Mitgliedschaftsmodell für wissenschaftliche Einrichtungen. Die Zusammenarbeit mit (nicht stimmberechtigten) Journalen gehört zum Finanzierungsmodell. Journale können einen „Subskriptionsplan“ für eigene Autoren abschließen. Es besteht eine technische und eine Entwicklungspartnerschaft mit der California Digital Library, bei der

Langzeitarchivierung mit DANS. Dryad verfügt über ein Team professioneller Kuratorinnen. Eingereichte Daten werden auf grundlegende technische, Lizenzen und Personenrechte betreffende Metadaten inhaltlich geprüft. Zum Angebot gehören Reviewer-File, Versionierung und Langzeitarchivierung. Erhoben werden gestaffelte non-profit Upload-Gebühren (Data Publishing Charge) nach Nutzergruppen oder aber Mitgliedsbeiträge. Der Download ist kostenfrei.

datadryad.org

16. EUDAT Collaborative Data Infrastructure (EUDAT CDI)

Eine sich an kollaborative Forschungsprojekte, kleinere Forschungseinrichtungen, die keine eigene Infrastruktur aufbauen können sowie an Infrastrukturen wie ESFRI richtende, selbst kollaborativ organisierte, kostenlose Datenplattform mit mehreren, Forschungsprozesse begleitenden Diensten sowie Anbindung an Hochleistungsrechner, die aktuell von 36 forschungsspezifischen als auch generischen Datenzentren angeboten wird. Die Dienste werden gemeinsam angeboten, der Betrieb liegt aber in der Verantwortung der jeweiligen Einrichtung. Die für den Zugang notwendige Registrierung nutzt akademische AAI-Methoden wie EduGain und ORCID. EUDAT ist Serviceprovider für EOSC und wird entsprechend von der EU als Projekt finanziert. Nach einer Registrierung ist ein unbegrenzter, nicht redigierter Upload möglich.

eudat.eu

17. Figshare

Ein kommerzielles, seit 2011 operatives, generisches Repositorium mit Sitz in UK und USA, das vom Technologieunternehmen Digital Science angeboten wird. Digital Science ist Teil der Verlagsgruppe Holtzbrinck (mit Sitz in Großbritannien),

zu der auch die Nature Publishing Group gehört. Einzelpersonen können kostenlos geringeren Speicher nutzen, Unternehmen wie z.B. Verlage oder gesamte Einrichtungen eigene institutionelle Instanzen aufbauen, die sie entweder komplett selbst betreiben oder durch Figshare hosten lassen können. Zum Kundenkreis gehören Journals (u.a. Wiley und Springer Nature) sowie Universitäten und Stiftungen (z.B. National Science Foundation, Health Research Alliance, Wellcome Trust). Technischer Dienstleister ist Amazon Web Services. Langzeitarchivierung sichert seit 2012 das CLOCKSS Netzwerk. Das Kostenmodell ist nicht transparent, Figshare behält sich außerdem vor, es jederzeit zu ändern.

figshare.com

18. Forschungsdatenzentrum der Bundesagentur für Arbeit (FDZ BA im IAB)

Das FDZ BA im Institut für Arbeitsmarkt- und Berufsforschung ist ein seit 2004 vom RatSWD akkreditiertes Forschungsdatenzentrum, das Forschenden weltweit einen Zugang zu administrativen und Befragungsdaten zu Personen, Haushalten und Betrieben in Deutschland ermöglicht. Die Metadaten sind frei zugänglich, Campus- und Public-Use-Files werden angeboten. Schwach anonymisierte Daten dürfen nur vor Ort, Scientific-Use-Files über Remote-Verbindung genutzt werden. Die Nutzung ist adressaten- und zweckgebunden; Verträge werden auf der Ebene von Institutionen geschlossen.

fdz.iab.de

19. Forschungsdatenzentrum der Statistischen Ämter des Bundes und der Länder (FDZ Bund, FDZ Länder)

Das FDZ Bund (seit 2001) und FDZ Länder (seit 2002) sind vom RatSWD akkreditiert und bieten Mikrodaten öffentlicher Hand für unabhängige

wissenschaftliche Forschung aus unterschiedlichen Bereichen (nichtwissenschaftliche Nutzung ist nicht gänzlich ausgeschlossen, muss aber gesondert beantragt werden). Institutionell finanziert und im Rahmen des Auftrags stellen die FDZ Bund und Länder nach einer verpflichtenden Schließung eines Nutzungsvertrags kostenpflichtig, zeitlich begrenzt, adressaten- und zweckgebundene Daten vor Ort oder über anonymisierte Scientific-Use-Files bereit. Zugang für Peer-Review ist möglich. Für die Lehre werden kostenfreie absolut anonymisierte Mikrodaten als Campus-Files angeboten. Umfangreiche Metadaten sind frei zugänglich.

forschungsdatenzentrum.de/de#datenangebot

20. Forschungsdatenzentrum des Leibniz-Instituts für Bildungsverläufe e.V. (FDZ-LifBi)

Das FDZ-LifBi stellt Erhebungsdaten des Nationalen Bildungspanels NEPS ab dem Jahr 2010 bereit. Die Arbeit rund um die föderal finanzierte Längsschnittstudie, in deren Rahmen regelmäßige Befragungen von Einzelpersonen erfolgen, ist seit 2014 als Leibniz-Institut für Bildungsverläufe e.V. verstetigt (An-Institut der Universität Bamberg). Zusätzliche Finanzierung erfolgt durch Drittmittelprojekte. Der kostenfreie Zugang ist streng geregelt, adressaten- und zweckgebunden, nur für Arbeitnehmerinnen wissenschaftlicher Einrichtungen nach Vertragsabschluss möglich und in verschiedenen Stufen geregelt: gering anonymisiert Daten vor Ort, moderat anonymisierte über ein biometrisches Authentifizierungssystem sowie stark anonymisierte Scientific-Use-Files zum Download. Das FDZ-LifB gehört zu den zehn Partnerinstitutionen des NFDI-Konsortiums KonsortSWD.

lifbi.de/Institut/Organisation/Forschungsdatenzentrum-LifBi

21. Forschungsdatenzentrum Sozio-oekonomisches Panel (FDZ SOEP)

Das Sozio-oekonomische Panel versteht sich als eine unabhängige forschungsbasierte Infrastruktureinrichtung. Aufbereitete und dokumentierte Mikrodaten aus der regelmäßigen Befragung von Einzelpersonen können von Forschenden weltweit kostenfrei genutzt werden (Längsschnittstudie). Der Zugang ist adressaten- und zweckgebunden, die Nutzungsverträge sind verpflichtend. Eine Nutzung ist vor Ort, über SOEPremote oder als Scientific-Use-File möglich. Für analytisch-didaktische Anwendungen in der universitären Lehre stellt das SOEP einen einfachen Stata-Übungsdatensatz im Panel-Format zur freien Verfügung bereit. 1983 als Teilprojekt des von der DFG geförderten SFB „Mikroanalytische Grundlagen der Gesellschaftspolitik“ entstanden, 1990–2002 als DFG-Projekt gefördert und seit 2000 mit einer Zusatzfinanzierung durch das BMBF ausgestattet ist das FDZ SOEP seit 2003 eine Serviceeinrichtung der Leibniz-Gemeinschaft, also aus Mitteln des BMBF und der Länder finanziert. Das FDZ SOEP gehört zu den zehn Partnerinstitutionen des NFDI-Konsortiums KonsortSWD.

diw.de/de/diw_01.c.600489.de/ueber_uns.html

22. Foto Marburg

Das 1913 gegründete Bildarchiv Foto Marburg (Deutsches Dokumentationszentrum für Kunstgeschichte) wird als Träger der Verbunddatenbank „Bildindex der Kunst und Architektur“ von der Universität Marburg überregional betrieben, aus zentralen Haushaltsmitteln, über Drittmittelprojekte sowie Nutzungsgebühren finanziert. Eine Abfrage sowie Downloads für wissenschaftliche Nutzung sind offen und kostenfrei, Verwertungsrechte bleiben dagegen beim Bildarchiv. Jegliche Verwendung des Bildmaterials zu Publikationszwecken ist ge-

nehmigungspflichtig. Kostenfreie Bildveröffentlichung gilt für Dissertationen, kunsthistorische und archäologische Fachzeitschriften sowie im Selbstverlag. Für andere wissenschaftliche Veröffentlichungen sowie für kommerzielle Vermarktung gelten gestaffelte Honorare. Weiterhin vermarktet Foto Marburg seine Bestände über Veröffentlichung von Mikrofiche-Indices in einem kommerziellen Verlag. Foto Marburg ist in das Portfolio des NFDI4Culture-Konsortiums eingebunden.

uni-marburg.de/de/fotomarburg

23. German Biobank Node

Die Biobanken der Universitätskliniken in Deutschland arbeiten im „German Biobank Node“ (GBN) zusammen. Zusätzlich hat sich mit zwei IT-Entwicklungszentren die German Biobank Alliance (GBA) zusammengeschlossen. Sie sind ein Teil vom BBMRI-ERIC (Biobanking and Biomolecular Resources Research Infrastructure – European Research Infrastructure Consortium), eine pan-europäische Infrastruktur nationaler Biobanknetzwerke. GBN vertritt darin die Interessen der deutschen Biobanken. Die Förderung liegt beim BMBF. Die Biobanken sammeln Bioproben als Datenspenden. Datensubjekte sind die potentiellen Spenderinnen – sie werden über Ziele, Vorgehen und Rechte ausführlich informiert. Sie erhalten keine persönlichen Vorteile. Allerdings wird in Aussicht gestellt, über „gesundheitsrelevante Erkenntnisse“ – sollten solche gewonnen werden – informiert zu werden. Hier wird also das „Bezahlen mit Wissen“ zur Motivierung der Datenspendenden eingesetzt. Biobanken verfügen sowohl über analoge (Probenmaterial) als auch digitale Daten. Über Anträge auf Zugang und Verwendung von Daten (und Proben) in einem Forschungsprojekt beraten zwei Gremien: ein Vergabegremium und eine Ethikkommission. Sie entscheiden, ob das Forschungsvorhaben geltenden Vorgaben

entspricht. Daten werden nur herausgegeben, wenn beide Gremien zustimmen.

bbmri.de

24. *Google Earth Engine*

Die cloudbasierte Analyseplattform sammelt offene Erdbeobachtungsdaten, stellt aktuelle und historische Satellitenbilder kostenlos für eine nicht-kommerzielle Nutzung (inkl. Data-mining) bereit. Eigene Daten können hinzugefügt werden. Zum Angebot gehören auch ein Visualisierungstool für einfache Analysen sowie eine Entwicklungsumgebung (Code Engine), die nach Erstellung eines Google-Accounts und Zulassung genutzt werden kann. Der Dienst wird aus allen Sektoren nachgefragt. Kommerzielle Nutzung wird lizenziert (Bedingungen nicht veröffentlicht). Der Betreiber selbst bezeichnet den Dienst als „education tool“.

earthengine.google.com

25. *Inorganic Crystal Structure Database (ICSD)*

Datenbank für Kristallografie mit über 240.000 Datensätzen, betrieben vom FIZ Karlsruhe (für Europa) und vom National Institute of Standards and Technology U.S. (NIST). Die Finanzierung ist teils durch institutionelle Förderung teils durch non-profit Gebühren gesichert. Zunächst am Institut für Anorganische Chemie der Universität Bonn und ab 1985 in Zusammenarbeit mit FIZ Karlsruhe betrieben, ging die Datenbank 1989 an ein Joint Venture zwischen dem Gmelin Institute Deutschland und FIZ Karlsruhe über. Von 1997 bis 2017 wurde ICSD gemeinsam von FIZ Karlsruhe und NIST, seit 2017 ausschließlich durch das FIZ Karlsruhe betrieben. Neben den eigenen Datensätzen werden Informationen aus Chemie, Physik, Mineralogie, Geologie und Materialwissenschaften für die wissenschaftlichen

und gewerblichen Nachfrager bereitgestellt. Gebühren sind nach Sektoren gestaffelt, für kommerzielle Nachfrager werden zusätzliche Dienstleistungen wie Durchführung von Auftragsanalysen angeboten. Die Datenbank wird durch eine Redaktion betreut (Editorial Team), mit halbjährlichen Updates (Wachstum ca. 7.000 Datensätze pro Jahr). Daten gehen zurück auf 1913. Die Datenbank soll sich künftig an ausgewählten Standards von NFDI4Chem orientieren.

icsd.products.fiz-karlsruhe.de

26. *Juris.de – Das Rechtsportal*

Datenbank mit juristischer Literatur (25 Millionen verlinkte Dokumente) aus amtlichen Quellen sowie Verlagen der jurisAllianz, die vorwiegend in der juristischen Praxis aber auch von der öffentlichen Verwaltung, hochschulischer Lehre und Unternehmen genutzt wird. Die Gründung geht auf einen Auftrag der Bundesregierung von 1973 zum Aufbau eines computergestützten Rechtsinformationssystems zurück. Die Ausgliederung führte zu einer Teilprivatisierung des Dienstes, der aktuell von einer GmbH mit Mehrheitsbeteiligung des Bundes sowie weiteren Beteiligungen eines französischen Verlages, der Bundesrechtsanwaltskammer sowie des Landes Saarland betrieben wird. Der Betrieb wird durch Entgelte auf Modulbasis finanziert. Die Recherche nach Bundesgesetzen ist kostenfrei.

juris.de/jportal/index.jsp

27. *Medizininformatik-Initiative (MII)*

MII richtet sich an universitätsmedizinische Standorte, an denen Daten von Forschung und Versorgung gesammelt werden. Die Datenintegrationszentren (DIZ) sind eng mit den klinischen Rechenzentren verbunden, womit eine Anbindung an die Systeme der Krankenversorgung gewährleistet ist. Die Daten wer-

den zusammengeführt und aufbereitet, dabei ist Datenschutz von zentraler Bedeutung. Die übergreifende Zusammenarbeit wird von einer Koordinationsstelle organisiert. Die Initiative wird vom BMBF gefördert.

medizininformatik-initiative.de/de/konsortien/daten-integrationszentren

28. Mendeley Data

Das fachübergreifende, seit 2015 operative Datenrepositorium wird von Elsevier betrieben und ist an ein durch den Verlag übernommenes soziales Forschernetzwerk gleichen Namens angedockt. Langzeitarchivierung erfolgt über DANS, als technischer Partner fungiert Amazon Web Services. Das Repositorium ist in andere modulare Angebote des Verlages wie z.B. LabBooks oder Data Management Tools sowie Reference Manager und Research Network Career eingebettet („Premium-Paket“). Datensätze unterliegen vor der Veröffentlichung einem Review-Verfahren durch die Betreiber, Zweitveröffentlichungen sind ausgeschlossen. Die angebotene Suche umfasst mehrere öffentlich zugängliche Datenrepositorien, so dass von außen keine Aussage über den hauseigenen Datenbestand möglich ist. Suche und Download erfordern keine Registrierung und sind kostenfrei. Datupload in kleinem Umfang ist für individuell Forschende ebenfalls kostenfrei. Angeboten werden Module für Institutionen und Journals anderer Verlage. Die Vertragskonditionen und Kostenmodelle sind nicht öffentlich.

data.mendeley.com

29. Mobilitätsdaten Marktplatz (MDM)

Der von der Bundesanstalt für Straßenwesen im Rahmen des Innovationsprogramms der Bundesregierung geförderte MDM soll den Austausch von Mobilitätsdaten befördern. An-

geboten wird zum einen eine „neutrale B2B-Plattform“ für ein sicheres und transparentes Anbieten, Recherchieren und Abonnieren von Verkehrsdaten für Projekte im Bereich automatisiertes und vernetztes Fahren, zum anderen die Vernetzung der Akteure: ein Themenforum, Konferenzen sowie eine User Group, die auch Empfehlungspapiere veröffentlicht. Neben der Wirtschaft wird auch der Wissenschaft die Nutzung des Datenangebots nahegelegt. MDM basiert auf einem europäischen Standard für den Datenaustausch im Bereich der Intelligen-ten Verkehrssysteme, der im Rahmen mehrerer aufeinander aufbauender Förderprogramme der EU-Kommission entwickelt wird.

mdm-portal.de

30. Novel Materials Discovery (NOMAD)

Zwischen 2015-2018 im Rahmen des Projekts „Novel Materials Discovery Centre of Excellence“ (Förderung im Rahmen der High Performance Computing Programmlinie von Horizont 2020, Industriebeteiligung) entstanden, umfasst das vom Fritz-Haber-Institut der MPG in Berlin koordinierte Konsortium acht europäisch, materialwissenschaftlich orientierte Forschungseinrichtungen und vier Supercomputer-Zentren – darunter Max Planck Computing and Data Facility (MPCDF). Zu dem Betrieb gehört ein kostenfreies Archiv, eine Materialien-Enzyklopädie (seit 2017) sowie Rechenkapazität und Software-Werkzeuge für Big Data Analytics. Daten und Algorithmen stehen frei zur Verfügung. Der Dienst befindet sich nach Auslaufen der Förderung in einer Übergangsphase und soll im NFDI-Konsortium FAIRmat weiter betrieben werden. Aktueller Träger ist FAIR-DI e.V. mit Sitz in Berlin. Der Dienst wird auch im EOSC-Kontext genannt.

nomad-coe.eu

31. *Ontosight*

Ontosight ist ein international aufgestelltes Unternehmen der Innoplexus AG, das Daten aus öffentlich finanzierten Infrastrukturen wie Patent- und Sequenzdatenbanken mit Informationen aus öffentlich zugänglichen Literaturdatenbanken verbindet. Den Datenerzeugern bietet der Dienst die Möglichkeit, bisher unveröffentlichte Daten zu monetarisieren. Die so kumulierten Daten können mithilfe einer selbstlernenden Ontologie durchsucht und mit bereitgestellten KI-Werkzeugen analysiert und visualisiert werden. Zusätzlich wird ein Newsletter zu ausgewählten Forschungsgebieten angeboten. Die Angebote sind nach Kundengruppen gestuft: (Pharma)Industrie, klinische Wirkstoffstudien und akademische Einrichtungen. Für die universitäre Forschung steht weltweit ein „Discover Academics Access Program“ kostenlos zur Verfügung.

ontosight.ai

32. *OpenJur*

Die von einem gemeinnützigen Verein betriebene Datenbank bietet seit 2009 Daten zur Rechtsprechung aus öffentlicher Hand, ergänzt um einen News- und Preeticker. Um den nachhaltigen Betrieb der Datenbank sicherzustellen, wurde 2020 die gemeinnützige Tochtergesellschaft openJur gUG (haftungsbeschränkt) gegründet, die sich der Initiative Transparente Zivilgesellschaft verpflichtet hat. Der openJur e.V. ist Alleingesellschafter. Registrierte Nutzer können eigene Daten beitragen. Diese werden vor der Veröffentlichung begutachtet. Die Inhalte sind offen und kostenfrei nutzbar, außer urheberrechtlich geschützter Abbildungen. Der Dienst wird aus Privatmitteln des Betreibers und Spenden finanziert. Der Betreiber engagiert sich darüber hinaus im Bereich Legal Tech, forscht im NLP-Bereich und nutzt dynamische Ver-

arbeitung von Inhalten, um sie auch für Dritte bereitstellen zu können.

openjur.de

33. *OpenStreetMap (OSM)*

Der 2004 etablierte Dienst bietet für vielfältige Nutzungsarten Karten unter der Open Database Licence an. Der Dienst wird kollaborativ verantwortet und über diverse Regionalvereine und eine weltweite Freiwilligengemeinde weiterentwickelt und kuratiert. Auch wissenschaftliche Einrichtungen unterstützen den Dienst: u.a. stellt eine britische Universität zentrale Server zur Verfügung. Die technische Infrastruktur wird aus Spenden finanziert. Die in Großbritannien ansässige OpenStreetmap Foundation organisiert zentrale Aktivitäten und Spendenakquise.

openstreetmap.de

34. *PANGAEA – Data Publisher for Earth & Environmental Science*

Ein zertifiziertes Repositorium für georeferenzierte Datensätze aus individueller Forschung und Forschungseinrichtungen. Es besteht eine enge Verbindung zum Publikationswesen: Datensätze können als Supplement zu Artikeln oder Datensammlungen in Kombination mit Data Journals der Domäne veröffentlicht werden. Die meisten Daten sind frei zugänglich, gegebenenfalls bestehen Lizenzbeschränkungen oder Moratorien für Daten aus laufenden Projekten. Datensätze werden wissenschaftlich kuratiert. Der Dienst wird vom Alfred-Wegener-Institut für Meeresforschung und dem Zentrum für Marine Umweltwissenschaften der Universität Bremen betrieben und ist mit Mitteln der DFG, des BMBF und der Europäischen Kommission co-finanziert. Die überregionale Verfügbarkeit des Dienstes ist nicht dauerhaft gesichert. Er

soll in der NFDI-Initiative NFDI4Earth und NFDI-4Biodiversity eine Rolle spielen.

pangaea.de

35. Prometheus – Das verteilte digitale Bildarchiv für Forschung und Lehre

Das von 2001-2004 vom BMBF als Verbundprojekt geförderte Bildarchiv erschließt in einer Datenbank (Meta-)Daten aus Gedächtniseinrichtungen. Der Dienst, der sich selbst als „Datenbankbroker“ bezeichnet, wird von einem gemeinnützigen Verein mit Sitz am Kunsthistorischen Institut der Universität zu Köln betrieben. Für den Bestand gelten jeweils Urheber- oder Verwertungsrechte der Geberinstitutionen. Der Dienst umfasst einen öffentlichen, gebührenfreien und einen lizenzpflichtigen Bereich. Nach erfolgreicher Registrierung können eigene Daten gespeichert und geteilt werden. Für Datennachfrager aus der Wissenschaft werden moderate, gestaffelte Campus- oder individuelle Lizenzen angeboten. Förderungen durch Stiftungen, Universitäten und BMBF ergänzen die Finanzierung. Betreiber und technischer Dienstleister ist die Universität Köln. Der Dienst spielt eine zentrale Rolle im NFDI-Konsortium NFDI4Culture.

prometheus-bildarchiv.de

36. PsychArchives – Disciplinary Repository for Psychological Science

Der Dienst wird vom Leibniz-Zentrum für Psychologische Information und Dokumentation im Rahmen eines institutionellen Auftrags entwickelt und betrieben. Er besteht aus einer fachspezifischen Publikationsplattform für begutachtete Forschungsdaten, Preprints, Primär- und Sekundärpublikationen sowie einem Suchportal, einer Data-Sharing-Plattform, Tools für TDM und maschinelles Lernen. Weitere Werkzeuge zur Studienplanung, Prä-Registrierung

psychologischer Studien, Datenerhebung und Datenanalyse werden entwickelt. Die Nutzung ist nach einer Registrierung kostenlos, Zugang zu den Inhalten ist durch Freigabestufen geregelt.

psycharchives.org

37. RADAR – Research Data Repository

Das generische Repositorium für Archivierung und Veröffentlichung von Forschungsdaten wird – nach einem von der DFG geförderten Aufbau – seit 2017 wissenschaftsintern vom FIZ Karlsruhe weiterentwickelt und betrieben. Es richtet sich an Datenerzeuger aus öffentlich geförderten Hochschulen und außeruniversitären Forschungseinrichtungen in Deutschland (insbesondere in zeitlich befristeten Projekten in Fächern, in denen noch keine disziplinspezifischen Lösungen zum Forschungsdatenmanagement existieren). Nach Vertragsabschluss durch institutionelle Kunden bzw. Projekte werden leistungsabhängig gestaffelte Nutzungsgebühren für nicht redigierte Uploads erhoben. Die Kosten für die Nutzung setzen sich aus einer Grundgebühr und einem nutzungsabhängigen Entgelt (je nach in Anspruch genommener Dienstleistung) zusammen. Zusätzlich zum Basis-Service (Archivierung, Veröffentlichung, Rollenmanagement, Peer-Review-Files, Metrics) sind erweiterte vertragliche Leistungen buchbar: Implementierung einer lokalen Instanz an der Einrichtung oder die Anpassung an das Corporate Design.

radar-service.eu/de

38. Socioeconomic Data and Applications Center (SEDAC)

Das Datenzentrum der NASA sammelt raumbezogene Daten zu Mensch-Umwelt-Interaktionen. Es umfasst eine kuratierte Datensammlung mit einer Sammlungsstrategie und einem begutachteten Auswahlprozess: Forschende können

ihren Datensatz einreichen, die Aufnahme erfolgt erst nach einer Begutachtung. Datensätze werden individuell lizenziert und können einzeln heruntergeladen werden. Zu den Daten sind oftmals Visualisierungen und Grafiken verfügbar. Der Service wird vom Center for International Earth Science Information Network betrieben, welches zur Columbia University gehört.

earthdata.nasa.gov/eosdis/daacs/sedac

39. *SowiDataNet | datorium*

Ein vom GESIS Leibniz-Institut für Sozialwissenschaften projektbasiert entwickeltes und institutionell betriebenes, fachspezifisches Forschungsdatenrepositorium für Sozial- und Wirtschaftswissenschaften, das ermöglicht, quantitative Primär- und Sekundärdaten aus individueller Forschung sowie aus Einrichtungen ohne eigene Infrastruktur zu archivieren bzw. zu veröffentlichen. Die Daten werden in vier Zugangskategorien eingeteilt, die von einem freien Zugang ohne Registrierung bis hin zur Embargofrist reichen. Die Zugangskategorie wird vom Datengebenden bestimmt. Das Angebot umfasst bitstream preservation; durchgeführt wird eine rudimentäre Überprüfung vor der Datenübernahme. Es wird durch die Anbindung an das Social Science Open Access Repository für Zweitveröffentlichungen ergänzt. Gebühren werden von den Datenerzeugern erhoben. Der Betreiber ist Sprecher des NFDI-Konsortiums KonsortSWD.

data.gesis.org/sharing/#!Home

40. *Statista*

Eine kommerzielle, weltweit tätige Datenplattform für statistische Daten, die sich selbst als eine Agentur für Recherche und Informationsdesign versteht. Das Geschäftsmodell basiert

auf der Aggregation von Daten aus 22.500 Quellen. Neben Datenbanken und Dossiers werden individualisierte Recherche- und Analyseleistungen angeboten. Der Dienst richtet seine Angebote (Freemium Modell, Campus-Lizenzen, Unternehmenslösungen) an Nachfrager aus der Wirtschaft, Wissenschaft und auch aus dem sekundären Bildungsbereich. Mehr als 2.000 Hochschulen erwerben Campus-Lizenzen für Forschung und Lehre. Der Zugang zu Basis-Statistiken ist kostenlos.

de.statista.com

41. *The Arabidopsis Information Resource (TAIR)*

Die Datenbank TAIR bietet Genom-Informationen und molekularbiologische Informationen zum pflanzlichen Modellorganismus *Arabidopsis thaliana*. Die Daten werden aus anderen Ressourcen übernommen und angereichert. Neben Literaturinformationen gibt es von der Betreiberfirma inhouse entwickelte Analyse- und Visualisierungstools. Der Dienst wurde zunächst primär über die Forschungsförderung in den USA aufgebaut. Danach übernahm ein non-profit Unternehmen (Phoenix Bioinformatics, eine Ausgründung der Projektmitarbeiter) den Betrieb. Die aufwendige Kuratierung wird sowohl durch den Betreiber wie auch durch die Nutzergemeinschaft geleistet. Datennachfragen kommen meist aus der Wissenschaft. Das Board of Directors ist mit Personen aus der Wissenschaft und aus dem Verlagswesen besetzt. Finanziert wird der Dienst durch moderate, jährliche Gebühren, für Einrichtungen gestaffelt nach Größe und Sektor. Für Nutzer aus Entwicklungsländern ist der Dienst kostenlos.

arabidopsis.org

42. *TriNetX*

Die 2013 als Start-up gegründete Plattform verknüpft Daten aus klinischen Studien mit Daten anderer Teilnehmer und ermöglicht ihre Analyse. Zusätzlich werden Data Warehouse, Cloud-Speicher, kollaboratives Arbeiten und Rechtemanagement angeboten. Kunden laden Daten hoch, behalten aber ihre Verfügungsrechte und entscheiden, wem Zugang gewährt wird. Das Unternehmen akquiriert Investorengelder und Aufträge für den Dienst. Mit Kliniken und Universitäten weltweit werden Einzelverträge abgeschlossen, die den Zugang zu Daten regeln.

trinetx.com

43. *World Stress Map (WSM)*

Der nach einer zwanzigjährigen Projektaufbauphase seit 2009 vom GFZ Potsdam verantwortete Dienst sammelt weltweit seismographische Daten in Zusammenarbeit zwischen öffentlicher Hand, öffentlich finanzierter Forschung und Industrieforschung. Das GFZ Potsdam ist Betreiber und Geldgeber. Die institutionelle Verantwortung ist dort dauerhaft angesiedelt. Für die Kuratierung wird ein „quality ranking scheme“ verwendet, das die weltweite Interoperabilität der Daten sichert. Das Datenschema wird von der Fachgemeinschaft mitbestimmt und setzt mittlerweile einen in der Disziplin international akzeptierten Standard. Die redaktionell kuratierte Datenbank ist aufgrund ihres Renommées Grundlage zahlreicher Forschungsprojekte, wird von öffentlicher Hand und von der Wirtschaft z.B. bei Vorbereitung von Bauvorhaben wie Atommüllendlager genutzt. Der Gesamtbestand ist für Nutzung und Verwertung aus allen Sektoren frei.

world-stress-map.org

44. *Worldwide Protein Data Bank (wwPDB)*

Die in den 1970er Jahren gegründete wwPDB ist eine transnationale Dateninfrastruktur, die nach mehrjähriger Förderung überwiegend durch die US-amerikanische National Science Foundation in ein internationales Konsortium aus vier Regionaldatenbanken überführt wurde, die von den für sie zuständigen Forschungsförderern finanziert werden. Die Daten werden kuratiert, stehen unter freier Lizenz kostenfrei zur Verfügung und werden in unterschiedlichen Sektoren nachgenutzt.

wwpdb.org

45. *Zenodo*

Ein generisches, seit 2015 bestehendes, bei Forschenden sehr populäres non-profit Repository für unredigierte, individuell lizenzierte Veröffentlichungen aller Art. Im Rahmen des EU-Projektes OpenAIRE co-finanziert, basiert der Dienst auf einer hausinternen Entwicklung des CERN und wird dort „on-top“ betrieben. Zusätzliche Förderung stammt von Stiftungen und in geringem Umfang aus akquirierten Spenden. Eine Kooperation mit Dryad mit dem Ziel der Verbesserung von Workflows und der Kuratierung besteht seit Februar 2021. Bisher ist nach einer durch ORCID oder EduGAIN autorisierten Registrierung ein unbegrenzter Upload möglich. Zenodo informiert ausdrücklich, dass keine Zusagen über dauerhafte Nachnutzbarkeit der Daten bestehen.

zenodo.org

B. DISKUSSIONSIMPULS: DATENDIENSTE NACHHALTIG GESTALTEN

Mai 2020

Datendienste nachhaltig gestalten

Ein Diskussionsimpuls zur Weiterentwicklung von Forschungsdateninfrastrukturen

Der Rat für Informationsinfrastrukturen (RfII) hat in den Jahren 2017 und 2018 Diskussionsimpulse zur Ausgestaltung einer Nationalen Forschungsdateninfrastruktur (NFDI) publiziert.¹ Mit der Bewilligung der ersten Konsortien werden ab 2020 Weichen für Datendienste gestellt, die dann in der Verantwortung der NFDI sowie von Bund und Ländern stehen werden. Aber auch außerhalb beziehungsweise parallel zur NFDI werden in Deutschland und in Europa Informationsinfrastrukturen weiterentwickelt und aufgebaut, unter anderem die *European Open Science Cloud* (EOSC) mit ihren transnationalen Diensten und Angeboten.

Die Gemeinsame Wissenschaftskonferenz (GWK) hat den RfII in der Rolle eines Beraters für Politik und Wissenschaft eingesetzt. Vor diesem Hintergrund wird er die Entwicklung im Bereich der Forschungsdateninfrastrukturen auch weiterhin durch Diskussionsimpulse kompakt und zeitnah unterstützen. Dabei ist ihm eine gut ausbalancierte Entwicklung wichtig, die sowohl die Überschneidung von Bedarfen von Wissenschaft, öffentlicher Verwaltung und Wirtschaft berücksichtigt als auch die internationale Zusammenarbeit.

NACHHALTIGES NUTZEN FÜR DIE WISSENSCHAFT ORGANISIEREN

Forschungsdatendienste in der öffentlichen Wissenschaft müssen darauf angelegt sein, eine wissenschaftsweite Nutzung in möglichst nachhaltiger Form zu gewährleisten. „Nutzung“ von Daten geht dabei zunächst nicht mit einer „Verwertung“ im kommerziellen Sinne einher.² Diese Zielstellung drückt sich zum Beispiel im breit akzeptierten Leitmotiv der Offenheit und Zugänglichkeit von Wissenschaft aus (mit einem breiten Spektrum an Programmatiken wie z. B. Open Access, Open Science oder FAIR). In Deutschland soll die NFDI vor eben diesem Hintergrund durch Communities beziehungsweise Fachgemeinschaften ausgestaltet werden, um sich dann, angetrieben durch die fachliche Nachfrage, als wissenschaftseigenes Netzwerk von Forschungsdatendiensten zu entwickeln. Darüber hinaus entstehen mit öffentlicher Förderung weitere Datensammlungen und -dienste für die Wissenschaft, die es anschlussfähig zu gestalten gilt.

¹ Vgl. RfII (2018) – Diskussionspapier: In der Breite und forschungsnah; RfII (2018) – Positionspapier Zusammenarbeit als Chance; RfII (2017) – Schritt für Schritt - oder: Was bringt wer mit?

² Nutzungen durch weitere Akteure im außerwissenschaftlichen Kontext sind nicht grundsätzlich ausgeschlossen, stehen aber bei der innerwissenschaftlichen Organisation von Daten meist nicht im Vordergrund.

Auf dem Weg zu guten Lösungen sind eine Vielzahl von Gestaltungsalternativen denkbar. Dies betrifft neben der Qualitätssicherung auch die „Betriebsmodelle“ und hierauf aufbauende „Geschäftsmodelle“ (in einem nicht profitorientierten, aber auf Nachhaltigkeit und Ressourcenschonung angelegten Wortsinn), die Verteilung von Aufgaben und Verantwortlichkeiten sowie die Bereitstellungs- und Nutzungsformen (mit ggf. Regeln für eine verantwortungsvolle Bereitstellung und Nutzung, Aneignungsverboten etc.). Der RfII formuliert vor diesem Hintergrund einige aus wissenschaftlicher Perspektive gebotene Anforderungen an Daten und Datendienste.

GESTALTUNGSMERKMALE FÜR DATEN(SAMMLUNGEN) UND DATENDIENSTE

Zielgruppenorientierung. Die Angebote müssen zunächst die Bedarfe ihrer (wissenschaftlichen) Zielgruppen reflektieren. Aushandlungsprozesse hierzu inklusiv und gerecht zu führen, gehört zu den grundlegenden Anforderungen zum Beispiel an die Konsortien der NFDI, aber auch an transnationale Strukturen wie die EOSC oder andere Forschungsdateninfrastrukturen.³ Während des Aufbaus sollte dabei die aktuell existierende wissenschaftliche Nachfrage, und nicht eine lediglich antizipierte (mögliche oder denkbare) künftige Nachfrage im Vordergrund stehen.

Standortneutrale Zugangsregime. Große nationale Forschungsdateninfrastrukturen wie die NFDI sind als überregionale, transdisziplinäre Netzwerke angelegt, mit dem Ziel eines gleichermaßen guten Zugangs für Wissenschaftlerinnen und Wissenschaftler unabhängig vom jeweiligen Standort. Sofern der Zugang für Nutzende an Voraussetzungen geknüpft ist (etwa an die Einwilligung in Datenschutzbestimmungen oder spezifische methodische Kompetenzen), sollten interne Regelungen sicherstellen, dass es keine regionalen oder organisationsbezogenen Bevorzugungen innerhalb und zwischen den Akteuren gibt, die das Netzwerk tragen. Wo Forschungsdateninfrastrukturen zunächst für den nationalen Bedarf konzipiert werden, müssen Anforderungen an die Anschlussfähigkeit zu supra- und internationalen Wissenschaftsstrukturen frühzeitig mitgedacht und mitgestaltet werden. Insbesondere die Kooperationsfähigkeit im europäischen Forschungsraum gilt es sicherzustellen.

Wissenschaftsweite Teilhabe. Daten und Datendienste sollten auch in fachlicher Hinsicht wissenschaftsweit und transdisziplinär nutzbar sein. Die in fachübergreifende Infrastrukturen wie zum Beispiel die NFDI ‚mitgebrachten‘ Metadaten müssen dort, wo dies möglich und sinnvoll ist, in gemeinschaftlich vereinbarte und idealerweise international anerkannte Ontologien überführt werden und sollten für den Einsatz maschineller Verfahren geeignet sein. Neben einer solchen Fortentwicklung der Aus- und Kennzeichnung von Forschungsda-

³ Vgl. zur Nutzereinbindung ausführlich: RfII (2018) – Diskussionspapier: In der Breite und forschungsnah, mit Blick auf europäische Gestaltungsansätze siehe RfII (2018) – Stellungnahme zur EOSC.

ten sollten auch Pfade für die Entwicklung transformierter Datensätze oder „Datenprodukte“ mitgedacht werden, die Nutzungsmöglichkeiten für andere Disziplinen und Domänen schaffen.⁴

Wissenschaftsadäquate Nutzungsbedingungen. Datendienste für die Forschung zeichnen sich durch die Qualitätsmerkmale der Nachvollziehbarkeit, der Beständigkeit und der Unabhängigkeit (von wissenschaftsfremden Interessen) aus. Nutzende sollten daher anbieterseitig nicht nur bestmögliche Informationen zur Provenienz und wissenschaftlichen Qualität von Daten erhalten, sondern es müssen bei Bedarf auch Langzeitverfügbarkeit von Inhalten, Zugangs- und Nutzungsrechten gewährleistet sein.⁵

Nutzung außerhalb der Wissenschaft. Wissenschaftseigene Angebote sollten grundsätzlich auch für Nachfragen aus Wirtschaft und Gesellschaft zugänglich sein. Gleichwohl können die Nutzungsinteressen und die Bedingungen der Arbeit mit den Daten dann andere sein als beim Teilen von Daten innerhalb der Wissenschaft. Aspekte der Datennutzung an der Schnittstelle zu Wirtschaft und Gesellschaft sollten deshalb beim Aufbau neuer Dienste konzeptionell und schon frühzeitig berücksichtigt werden. Bereits praktizierte und erprobte Modelle sowie Erfahrungswerte aus der Breite der beteiligten wissenschaftlichen Communities und Fachgemeinschaften können hierfür eine Grundlage bieten.

Nutzung und qualitative Erweiterung von Daten im Forschungsprozess. Datendienste sollen als eine gemeinsame Ressource den Fortschritt von Forschung steigern können. Deshalb ist es erstrebenswert, dass Forschungsergebnisse, die beispielweise mit Daten und Datenprodukten aus wissenschaftlichen Beständen erzielt wurden – etwa erweiterte oder transformierte Daten – auch wieder in den Bestand zurückgeführt werden und so zur Erweiterung und Qualität des Datenpools beitragen. Um dieses Ziel zu erreichen, sollten die Betreiber der wissenschaftlichen Datendienste geeignete Vereinbarungen mit den Datennachfragern und -nutzern treffen.⁶

Wissenschaftliche Reputation. Dateninfrastrukturen und ihre Dienste bringen die Forschung in dem Maße entscheidend voran, wo (und wie gut) sie die Qualität von Forschungsdaten, Methodik und Forschungsergebnissen sichern. Neben den auf diese Weise zu erzielenden Reputationsgewinnen sollten auch die Vorbereitung von Daten zur wissenschaftlichen Nutzung und die Entwicklung sowie der Betrieb von Datendiensten in geeigneter

⁴ Eine domänenübergreifende Erschließung von Daten kann mit Aufwand verbunden sein, vgl. RfII (2019) – Herausforderung Datenqualität, S. 75f., und dort zitierte Literatur. Dies gehört zu den Gründen, warum die Idee einer „offenen Bereitstellung“ im konkreten Fall fachliche Anstrengungen erfordern kann, um die erwünschte Nachnutzung zu gewährleisten (vgl. RfII (2019) – Stellungnahme aktuelle Entwicklungen Open Data).

⁵ Forschungsdateninfrastrukturen bzw. ihre Betreiber sollten insbesondere Aussagen zur Betriebskontinuität der Dienste treffen. Nach Recherchen des RfII befinden sich Nutzungsbedingungen und Dienstverträge im Bereich wissenschaftlicher Daten bislang oftmals nicht auf dem Stand von Regelungen, die z. B. für Zugänge im Bereich der digitalen wissenschaftlichen Literatur üblich sind.

⁶ Siehe hierzu auch: RfII (2018) – Diskussionspapier: In der Breite und forschungsnah, S. 5.

Weise reputationsförderlich sein.⁷ Die im Aufbau befindliche NFDI kann hierzu einen wesentlichen Beitrag leisten und auch entsprechende Impulse in den europäischen Forschungsraum geben.

Kompetenzaufbau. Nach wie vor beobachtet der RfII starke Engpässe und fehlende Qualifizierungen für Fachkräfte an der Schnittstelle Wissenschaft/Informationsinfrastruktur. Forschungsinfrastrukturen kommt hier eine besondere Rolle zu.⁸ Neben der Vermittlung von fachspezifischen, technischen und Informatikbezogenen Qualifikationen werden im Bereich der Koordinierung und Steuerung von Infrastrukturverbänden und Datendiensten auch wirtschaftliche und juristische Kompetenzen erworben, zum Beispiel bei der Ausarbeitung von Betriebs- und Geschäftsmodellen oder der Vertragsgestaltung mit kommerziellen Dienstleistern beziehungsweise Anbietern. Insbesondere die Konsortien der NFDI sollten sich in dieser Hinsicht auch als Treiber für die Heranbildung von Fachkräften begreifen.

Technisch-organisatorische Anschlussfähigkeit. Zum Umgang mit und zur Nutzung von Referenzmodellen sollte unter den Betreibern wissenschaftlicher Datendienste ein breiter Diskurs begonnen werden.⁹ Dies gilt in besonderem Maße für die NFDI-Konsortien, die auch existierende respektive parallel entstehende wissenschaftliche Informations- beziehungsweise IT-Infrastrukturen sowie hierauf zugeschnittene Betriebsmodelle im Blick behalten müssen: etwa auf lokaler Ebene der Universitäten und Forschungseinrichtungen, aber auch im Bereich des Hoch- und Höchstleistungsrechnens und der EOSC. Ein Zusammenspiel mit Bestrebungen in den Sektoren Wirtschaft und Verwaltung, Datenbestände zu vernetzen, wäre darüber hinaus förderlich.

Reform- und Kompromissbereitschaft. Zur Bildung fach- und domänenübergreifender Forschungsdateninfrastrukturen muss es zu Abstimmungen und gemeinsamen Entscheidungen zwischen den Akteuren kommen, die technische Systeme, verwendete Standards, gemeinsam zu nutzende Dienste, die Übernahme von Kosten oder den Austausch von Know-how und gegebenenfalls Ressourcen betreffen. Dies kann in Teilbereichen bedeuten, dass man etablierte Verfahren aufgeben oder anpassen muss. Hier sind zugunsten übergeordneter Ziele und der Finanzierbarkeit des Gesamtsystems Kompromissfähigkeit und eine Kommunikationskultur geboten, die Partikularinteressen überwinden hilft.

Prinzip der besten Eignung. Die oftmals gemeinschaftliche Organisation von Diensten bringt Arbeits- beziehungsweise Aufgabenteilungen mit sich, die sowohl zwischen den innerwis-

⁷ Der RfII hat sich zur Notwendigkeit einer Steigerung der Reputation von wissenschaftlicher Datenarbeit und der Erstellung wissenschaftlicher Datenprodukte in seinem Positionspapier „Herausforderung Datenqualität“ geäußert, RfII (2019) – Herausforderung Datenqualität.

⁸ RfII (2019) – Digitale Kompetenzen.

⁹ Zu nennen wäre etwa die ausdifferenzierte Referenzarchitektur für Datenplattformen der International Data Spaces Association (Whitepaper 2019) oder das OAIS Referenzmodell für die Langzeitarchivierung. Auch Reifegradmodelle für Forschungsinfrastrukturen (z. B. RISE, RISE-DE) sowie Standardklauseln für wissenschaftsadäquate Nutzungsbedingungen erscheinen sinnvoll.

senschaftlichen Verbundpartnern auszuhandeln sind, als auch mit externen, wissenschaftlichen sowie kommerziellen Akteuren und ihren zu integrierenden Angeboten. Handlungsleitendes Prinzip sollte dabei – nebst der Wahrung der Souveränität der Wissenschaft – die Übertragung einer Aufgabe an den jeweils fachlich-technisch am besten geeigneten Akteur sein. Grundlage ist eine klare Beschreibung der Aufgaben und ein Selbstverständnis darüber, welche Verantwortlichkeiten in der Hand der Wissenschaft verbleiben und welche Aufgaben in Zusammenarbeit mit externen Dienstleistern oder Partnern gut, international anschlussfähig und vor allem wissenschaftskonform gelöst werden können.

„**Bezahlen mit Daten**“. Beim Aufbau eines Dienste-Portfolios werden Entscheidungen zur Einbindung von Drittanbietern getroffen werden müssen. Hierbei ist Sorge zu tragen, dass Geschäftsmodelle Dritter, die Daten und Datenspuren aus wissenschaftlichen Arbeitsprozessen beziehungsweise Zwischenergebnisse aus der Forschung aufzeichnen und kommerziell vermarkten, nicht zulasten der Rechte der einzelnen Wissenschaftlerinnen und Wissenschaftler, der Forschung insgesamt oder der Ziele von wissenschaftlichen Einrichtungen und Infrastrukturverbänden gehen.¹⁰ Im Einzelfall stellen sich Grenzfragen, die wissenschaftsintern noch nicht hinreichend ausgeleuchtet oder gar ausverhandelt sind. Hierauf abzielende Verständigungsprozesse sollten im Gesamtsystem von den Verantwortlichen möglichst transparent und auch kritisch geführt werden.

Datensouveränität der Forschenden. Wo öffentlich geförderte wissenschaftliche Akteure beziehungsweise von ihnen gebildete Verbundstrukturen oder Einrichtungen der Infrastruktur Verträge mit (kommerziellen) Intermediären schließen, ist die Wahrung der Datensouveränität der Forschenden sowie der Wissenschaft und ihrer Institutionen insgesamt eine wichtige Anforderung an die Vertragsgestaltung, auch im Detail. Entsprechende Verträge mit wirtschaftlichen Partnern sollten öffentlich zugänglich sein und in einer Weise ausgestaltet werden, die einer für die Wissenschaft nachteiligen Monopolbildung entgegenwirkt.¹¹

10 Zu „Daten als Gegenleistung“ vgl. auch Datenethikkommission (2019) – Gutachten der Datenethikkommission.

11 So wie es auch die Eckpunkte einer Datenstrategie der Bundesregierung vorsehen. Siehe unter <https://www.bundesregierung.de/re-source/blob/975226/1693626/60b196d5861f71cdefb9e254f5382a62/2019-11-18-pdf-datenstrategie-data.pdf?download=1> (zuletzt geprüft am 04.05.2020).

C. MITWIRKENDE

MITGLIEDER DES RATES (STAND: JUNI 2021)

Vertreter der wissenschaftlichen Nutzer

Prof. Dr. Marion Albers

Universität Hamburg – Juristische Fakultät

Prof. Dr. Lars Bernard (stellv. Vorsitzender)

Technische Universität Dresden – Umweltwissenschaften

Prof. Dr. Stefan Decker

FIT – Fraunhofer-Institut für Angewandte Informationstechnik und RWTH Aachen

Prof. Dr. Petra Gehring (Vorsitzende)

Technische Universität Darmstadt – Institut für Philosophie

Prof. Dr. Kurt Kremer

MPI – Max-Planck-Institut für Polymerforschung Mainz

Prof. Dr. Wolfgang Marquardt

Forschungszentrum Jülich GmbH

Prof. Dr. Stefanie Speidel

Nationales Centrum für Tumorerkrankungen Dresden (NCT/UCC)

Prof. Dr. Joachim Wambsganß

ZAH – Zentrum für Astronomie der Universität Heidelberg

Vertreter von Bund und Ländern

Rüdiger Eichel

Niedersächsisches Ministerium für Wissenschaft und Kultur

Dr. Hans-Josef Linkens

Bundesministerium für Bildung und Forschung

Dr. Dietrich Nelle

Bundesministerium für Bildung und Forschung

Annette Storsberg

Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen

Vertreter des öffentlichen Lebens

Dr. Anke Beck

Foresight GmbH

Marit Hansen

Landesbeauftragte für Datenschutz Schleswig-Holstein

Christine Regitz

SAP SE

Dr. Harald Schöning

Software AG

Vertreter der Einrichtungen

Sabine Brünger-Weilandt

FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur GmbH

Prof. Dr. Barbara Helwing

Vorderasiatisches Museum Berlin – SMB SPK

Prof. Dr. Michael Jäckel

Universität Trier

Prof. Dr. Stefan Liebig (stellv. Vorsitzender)

DIW – Deutsches Institut für Wirtschaftsforschung e.V.

Prof. Dr. Sandra Richter

Deutsches Literaturarchiv Marbach

Katrin Stump

Universitätsbibliothek Braunschweig

Prof. Dr. Klaus Tochtermann

ZBW – Leibniz-Informationszentrum Wirtschaft

Prof. Dr. Ramin Yahyapour

GWdG – Gesellschaft für Wissenschaftliche Datenverarbeitung mbH Göttingen

PROJEKT NUTZUNG UND VERWERTUNG

Arbeitsgruppe Nutzung und Verwertung

Prof. Dr. Petra Gehring (Leitung), Prof. Dr. Marion Albers, Dr. Anke Beck, Prof. Dr. Lars Bernard, Sabine Brünger-Weilandt, Prof. Dr. Stefan Decker, Prof. Dr. Michael Jäckel, Prof. Dr. Stefan Liebig, Prof. Dr. Wolfgang Marquardt, Dr. Dietrich Nelle, Dr. Harald Schöning, Katrin Stump

Gäste der Arbeitsgruppe

Marit Hansen, Dr. Nicola Jentzsch (als Mitglieder der AG „Datentreuhänderschaft“ in ausgewählten Sitzungen)

Redaktionsgruppe

Prof. Dr. Petra Gehring (Leitung), Dr. Anke Beck, Katrin Stump und Prof. Dr. Joachim Wambsganß

Gremienbetreuung

Die Gremien wurden seitens der RfII-Geschäftsstelle inhaltlich und organisatorisch begleitet durch Dr. Stefan Lange und Dr. Beata Mache sowie durch Dr. Barbara Ebert (Arbeitsgruppe).

IMPRESSUM

Verabschiedet im Juni 2021

Rat für Informationsinfrastrukturen (RfII)
Geschäftsstelle
Papendiek 16
37073 Göttingen

Tel. 0551-392 70 50
E-Mail info@rfii.de
Web www.rfii.de

GESTALTUNG, SATZ UND DRUCK
NEFFO DESIGN (Buchholz), Klartext GmbH (Göttingen)

ZITIERVORSCHLAG

RfII – Rat für Informationsinfrastrukturen: Nutzung und Verwertung von Daten im wissenschaftlichen Raum – Empfehlungen zur Ausgestaltung von Datendiensten an der Schnittstelle zwischen Wissenschaft und Wirtschaft, Göttingen 2021, 120 S.

Der RfII bevorzugt eine gendergerechte Sprache. Es werden auch Kollektivbezeichnungen verwendet, die jeweils Personen aller Geschlechter einbeziehen.

Dieses Werk ist lizenziert unter einer  Creative Commons Namensnennung – Weitergabe unter gleichen Bedingungen 4.0 International (CC BY-SA 4.0).



Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über URN: [urn:nbn:de:101:1-2020052673](https://nbn-resolving.org/urn:nbn:de:101:1-2020052673) abrufbar.

