

Peter Schirnbacher

# Dimensionen des Forschungsdatenmanagements im digitalen Zeitalter

**Abstract:** Forschungsdaten als Ausgangspunkt und Ergebnis der wissenschaftlichen Arbeit gibt es, seit der Mensch Forschungen betreibt. Das Besondere der gegenwärtigen Situation ist, dass diese Forschungsdaten meistens digital vorliegen, sie in ihrem Umfang beträchtliche Ausmaße angenommen haben und uns in deutlich komplexerer Gestalt begegnen, als wir es von früher kannten. Der Artikel befasst sich mit den notwendigen Veränderungen des Datenmanagements im digitalen Zeitalter und systematisiert die Vielzahl von neuen Fragestellungen in den drei Dimensionen des Forschungsdatenmanagements: wissenschaftspolitisch, organisatorisch, technisch. Dabei werden deutlich mehr Fragen aufgeworfen als beantwortet, weil zum einen der gegenwärtige Erkenntnisstand noch sehr begrenzt ist und zum anderen eine größere Ausführlichkeit den Rahmen des Beitrages sprengen würde.

## 1 Einleitung

Es ist nicht zweifelsfrei nachweisbar, wer denn nun im deutschsprachigen Raum zuerst den Slogan: „Daten sind der Rohstoff des 21. Jahrhunderts“ benutzte. Die Äußerungen, jeweils im Umfeld der BitKom 2015 (Digitale Weiterbildung 2015) und der CeBit 2016 (Bundeskanzlerin Merkel 2016) getätigt, sollten ein zusätzlicher Anstoß zur Auseinandersetzung mit den sogenannten Big Data initiieren. Auf jeden Fall war es ein verspätetes Signal für Deutschland, den Umgang mit Daten noch mehr in den Mittelpunkt der Diskussionen sowohl in der Wirtschaft als auch in der Wissenschaft zu stellen als schon zuvor.

Big Data ist ein Buzzword der neueren Zeit, ohne das Informatiker und Informationswissenschaftler meinen, nicht mehr auskommen zu können. Letztlich soll zum Ausdruck gebracht werden, dass die Menge der Daten in der heutigen Zeit überdurchschnittlich steigt, zusätzlich durch eine besonders hohe Komplexität charakterisiert ist, als wenig strukturiert angesehen wird und in vielen Fällen sich durch Schnellebigkeit beschreiben lässt (Big Data 2013).

Eng verknüpft mit dem Begriff Big Data ist die digitale Erscheinungsform. Erst dadurch, dass die Daten digital vorliegen, lassen sich so extrem große Mengen an Daten durch den Menschen unter Zuhilfenahme von Computern

verwalten und ihre Komplexität beherrschen. Wir befinden uns also nach wie vor im Übergang vom analogen in das digitale Zeitalter, wobei ausdrücklich betont sei, dass damit die analoge „Welt“ nicht ersetzt wird, sondern eine Ergänzung und Erweiterung erfährt.

Betrachtet man diese vier Merkmale von Big Data als gegeben, so lässt sich unschwer erkennen, dass dies ein Phänomen in nahezu jedem gesellschaftlichen Bereich ist. In dem vorliegenden Beitrag soll es ausschließlich um Daten im Umfeld von Forschung und Wissenschaft gehen. Hier hat sich seit geraumer Zeit der Begriff der Forschungsdaten etabliert, wobei man die Forschungsdaten nicht unbedingt als eine 100%ige Teilmenge von Big Data ansehen, sondern vielmehr ihre Spezifik im Sinne des Entstehungs- und Verwertungsprozesses betrachten sollte.

„Unter digitalen Forschungsdaten verstehen wir alle digital vorliegenden Daten, die während des Forschungsprozesses entstehen oder ihr Ergebnis sind. Der Forschungsprozess umfasst dabei den gesamten Kreislauf von der Forschungsdatengenerierung, z. B. durch ein Experiment in den Naturwissenschaften, eine dokumentierte Beobachtung in einer Kulturwissenschaft oder eine empirische Studie in den Sozialwissenschaften, über die Bearbeitung und Analyse bis hin zur Publikation und Archivierung von Forschungsdaten. Digitale Forschungsdaten entstehen in allen Wissenschaftsdisziplinen und unter Anwendung verschiedener Methoden, abhängig von der Forschungsfrage. Dies hat zur Folge, dass sie in unterschiedlichen Medientypen, Aggregationsstufen und Datenformaten auftreten.“ (Kindling & Schirmbacher 2013)

Forschungsdaten sind äußerst heterogen, weil sie sich spezifisch für das jeweilige Wissenschaftsgebiet darstellen. Das macht den Umgang mit den Forschungsdaten deutlich komplizierter und komplexer und fordert disziplinspezifische Lösungen.

Forschungsdaten sind jedoch in keiner Weise etwas Neues, auch wenn man heute manchmal beim Blick in die Literatur den Eindruck gewinnen könnte, als würde das Rad zum ersten Mal rund sein. Forschungsdaten waren seit jeher der Ausgangspunkt von Forschung und ebenso ein Resultat der Forschung. Der wesentliche Unterschied und damit die „Berechtigung“, sich mit dem Thema Management von Forschungsdaten gegenwärtig intensiv zu befassen, liegt vorrangig in der Tatsache begründet, dass wir es heute in der überwiegenden Zahl der Fälle nicht mit analogen Daten, sondern mit digitalen Daten zu tun haben. Hier fehlt es der Community jedoch mindestens in zweierlei Hinsicht an Erfahrung. Zum einen gibt es für den Umgang mit digital vorliegenden Daten nur unzureichende oder gar keine Standards und kein passendes Regelwerk für die Speicherung, die Aufbewahrung, das Wiederauffinden und das Bereitstellen der Daten und zum anderen stehen wir vor demselben Phänomen wie beim Management von Big Data. Weil die technischen Gegebenheiten

es ermöglichen, haben wir deutlich mehr Daten mit wesentlich komplexeren Strukturen, die es zu verwalten gilt.

Die gegenwärtige weltweite intensive Beschäftigung mit digitalen Forschungsdaten ist in erster Linie der Tatsache geschuldet, dass die aus der Vergangenheit gewohnte Informationsinfrastruktur sich den Herausforderungen der digitalen Welt stellen muss, sich anpassen bzw. sich teilweise neu gestalten. Wie diese veränderte Form aussehen soll, wirft eine Vielzahl von Fragen auf. Der Beitrag versucht sie zu systematisieren. Dabei werden deutlich mehr Probleme aufgezeigt als Lösungsansätze dargestellt und diskutiert werden können. Das liegt zum einen am gegenwärtig noch begrenzten Erkenntnisstand und zum anderen am begrenzten Umfang eines Beitrages für einen Sammelband.

## 2 Dimensionen des Forschungsdatenmanagements

### 2.1 Vorbemerkungen

In einer größeren Zahl von Artikeln werden Aktivitäten der unterschiedlichsten Art von den jeweiligen Akteuren des Forschungsdatenmanagements (FDM) gefordert und so kommen Planungen, Projekte und Konzepte mit einer immensen Spannweite ans Tageslicht. Auf der einen Seite ist es die Absicht, einen weltumspannenden Service aufzubauen, und auf der anderen Seite geht es lediglich um den notwendigen Speicherplatz in einem universitären Rechenzentrum (Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz. Kommission Zukunft der Informationsinfrastruktur 2011; Humboldt-Universität zu Berlin 2014b; Van der Graaf & Waaijers 2012).

Auch der Wissenschaftsrat der Bundesrepublik hat sich mit dieser Thematik befasst und seine Empfehlungen „Zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020“ nach Empfängergruppen spezifiziert (2012, S. 10 ff.), die sich an einen Adressatenkreis richten, angefangen beim Bund und den Länderministerien bis hin zu den Wissenschaftlern. Dabei haben die aufgeworfenen Fragestellungen und Forderungen (hier natürlich als Empfehlungen formuliert) sehr unterschiedliche Tragweite. In der Soziologie würde man in diesem Zusammenhang von unterschiedlichen Ebenen der Auseinandersetzung sprechen und damit eine Makro-, Mikro- bzw. Mesoebene unterstellen. Das passt für die Organisation des FDMs jedoch nur zum Teil. Im Wesentlichen sind drei Gruppen von Fragestellungen zu sehen.

Das sind erstens Fragen, die einer wissenschaftspolitischen oder gesellschaftspolitischen Beantwortung bedürfen und somit sicher auf der Makroebene anzusiedeln wären. Zweitens sind es Fragen, die eine Verständigung verschiedener (Service-)Organisationen notwendig machen, um Zuständigkeiten und Verantwortlichkeiten zu regeln. Diese Fragegruppe könnte man der Mesoebene zuordnen. Nur begrenzt passfähig wäre die Benutzung des Begriffes Mikroebene für die Gruppe von Fragen, die sich mit technischen Standards und Verfahren beschäftigen. Gemeinhin unterstellen die Soziologen, dass es sich bei der Mikroebene um die Beschäftigung mit dem Individuum und seinen unmittelbaren Wechselwirkungen zu anderen Individuen bzw. zur Umgebung handelt, wovon man beim Forschungsdatenmanagement zumindest nicht unmittelbar sprechen kann. Unverfänglicher ist da sicher die Nutzung des Begriffes „Dimension“.

## 2.2 Wissenschaftspolitische Dimension des Forschungsdatenmanagements

„Our vision is a scientific e-infrastructure that supports seamless access, use, re-use, and trust of data. In a sense, the physical and technical infrastructure becomes invisible and the data themselves become the infrastructure – a valuable asset, on which science, technology, the economy and society can advance.“ (High Level Expert Group on Scientific Data 2010)

Über die grundsätzliche Zielstellung, wie sie hier in einer durch die EU in Auftrag gegebenen Studie geäußert wird, besteht in der einschlägigen nationalen und internationalen Literatur Einigkeit (National Science Foundation, o.J.; Joint Information System Committee 2016; Australian Research Council 2015). Die Kommission Zukunft der Informationsinfrastruktur (KII) geht in ihrer Schrift „Gesamtkonzept für die Informationsinfrastruktur in Deutschland“ vom April 2011 jedoch noch weiter:

„Gegenwärtig steigt die Menge der Forschungsdaten dramatisch an, die durch aufwändige Erhebungsverfahren und meist unter Einsatz erheblicher öffentlicher Mittel gewonnen werden. Sie sind als nationales Kulturgut anzusehen und sollten im Sinne einer hoheitlichen Aufgabe dauerhaft gesichert und der (Fach)Öffentlichkeit sowie zukünftigen (Forscher)Generationen zur Nachnutzung bereitgestellt werden.“ (Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz. Kommission Zukunft der Informationsinfrastruktur 2011, S. B109)

Für die Kommission sind die Forschungsdaten als ein nationales Kulturgut anzusehen, dessen Sicherung eine hoheitliche Aufgabe des Staates darstellt. Die-

ser Ansicht kann man sich im Wesentlichen anschließen, wenn man den Umgang mit Forschungsdaten in etwa gleichsetzt mit dem Stellenwert einer wissenschaftlichen Publikation. Dabei sei ohne Zweifel anerkannt, dass der entscheidende Output des Forschungsprozesses nach wie vor in eine Veröffentlichung mündet. Die Bedeutung der Aufbewahrung und Bereitstellung von Forschungsdaten nimmt jedoch zu. Erkennt man das an, ergibt sich daraus eine wissenschaftspolitische Aufgabenstellung, deren Bewältigung grundsätzlicher Entscheidungen des Staates, der europäischen Kommission und darüber hinaus bedarf. Der Handlungsspielraum liegt in diesem Fall also vom Grundsatz her außerhalb der wissenschaftlichen Community, was als wesentliches Charakteristikum für die wissenschaftspolitische Dimension des Managements von Forschungsdaten angesehen werden kann. Es geht um die Gestaltung der gesellschaftlichen Rahmenbedingungen. Naturgemäß sind diese vorrangig politisch geprägt oder zumindest nur eingeschränkt durch die Forschenden selbst beeinflussbar. Im Wesentlichen lassen sich drei Problemkreise unterscheiden: Gestaltung einer Grundstruktur für das FDM, Schaffung der rechtlichen Rahmenbedingungen und Absicherung der finanziellen Notwendigkeiten.

### **2.2.1 Gestaltung einer Grundstruktur für das Forschungsdatenmanagement**

Die entscheidende Frage ist in diesem Zusammenhang, ob die Wissenschaftspolitiker von Bund und Ländern ebenso bereit sind, Forschungsdaten als nationales Kulturgut zu verstehen und die daraus folgenden Konsequenzen entsprechend zu tragen. Der auf Empfehlung sowohl von KII als auch des Wissenschaftsrates (2012) im Herbst 2014 gebildete Rat für Informationsinfrastruktur (RfII) bekennt sich zwar nicht ausdrücklich dazu, lässt jedoch in seinem ersten Empfehlungspapier „Leistung aus Vielfalt“ (2016) keine Zweifel, diese Frage positiv zu beantworten. Er schlägt vor, eine nationale Forschungsdateninfrastruktur (NFDI) aufzubauen und schreibt dazu:

„Diese Nationale Forschungsdateninfrastruktur (NFDI) sollte die Form eines Netzwerkes haben, disziplinen- bzw. communityübergreifend angelegt sein und sowohl die existierenden großen Informationseinrichtungen als auch die nationale Ebene der ESFRI-Projekte und die Repositorien weiterer, in ihren Bedarfen hinreichend homogener Nutzergruppen einbinden. Als Verbund muss die NFDI eine auf Handlungsfähigkeit und die Steuerung von Entwicklungen (hier zunächst: eines Transitionsprozesses) ausgelegte Governance besitzen.“ (Rat für Informationsinfrastrukturen 2016, S. 40)

Festzulegen sind auf dieser Ebene der gestaltenden Entscheidungsprozesse Verantwortlichkeiten im Sinne von möglichen Sammelaufträgen und Verfahrensweisen des Zusammenwirkens von Ländern, Bund, EU und darüber hi-

naus. Der RfII nennt dies die Spannungsfelder in einem lernenden System mit verschiedenen Akteuren und führt Punkte an wie (Rat für Informationsinfrastrukturen 2016, S. 11):

- „Projektförderung vs. Institutionalisierung,
- Top down vs. bottom up koordinierte Prozesse,
- Open Data vs. Datensouveränität zur Eigenforschung (Kooperation vs. Konkurrenz),
- Disziplinäre bzw. institutionenspezifische Lösungen vs. übergreifende Dienste,
- Datenschutz vs. Forschungsfreiheit,
- Staatliche Förderung vs. privatwirtschaftliches Engagement,
- Infrastrukturleistungen behördlicher Art vs. Angebote von Großforschungs- oder Hochschuleinrichtungen,
- Investitionen in Infrastrukturen vs. Investitionen in Köpfe,
- Nationale Aktivitäten vs. europäisches und globales Umfeld“.

Die wissenschaftspolitische Dimension beschränkt sich nicht ausschließlich auf Länder- Bundes- oder EU-Ebene, sondern umfasst in gleicher Weise Fragestellungen, die sich auf Hochschulleitungen bzw. Leitungen von außeruniversitären Forschungseinrichtungen beziehen. So hat sich insbesondere die Kommission „Digitale Infrastrukturen“ der Hochschulrektorenkonferenz (HRK) sehr intensiv mit dem Thema FDM auseinandergesetzt und entsprechende Empfehlungen für die Hochschulen erarbeitet (Hochschulrektorenkonferenz 2014; Hochschulrektorenkonferenz 2015). Sechs Schwerpunkte stehen dabei im Mittelpunkt:

- Orientierung geben,
- Datenkultur stärken und Anreize schaffen,
- Strategie entwickeln,
- Umsetzung organisieren,
- Infrastrukturen ausbauen,
- Kompetenzen weiterentwickeln.

Die drei ersten Punkte tragen dabei strategischen Charakter und bedürfen des tatkräftigen Einsatzes der jeweiligen Hochschulleitungen, wobei die drei dann folgenden Punkte etwas mehr die gestalterischen Komponenten unterstreichen und somit mehr der organisatorischen Dimension anzurechnen wären.

### **2.2.2 Schaffung der rechtlichen Rahmenbedingungen**

Außerhalb der unmittelbaren Einflussmöglichkeiten der Forschenden liegt eine Vielzahl von rechtlichen Fragen. Sie betreffen in erster Linie das

Urheberrecht, das Leistungsschutzrecht, das Vertragsrecht und das Wissenschaftsrecht. Notwendig ist eine gründliche Bearbeitung unter nationalen und internationalen Gesichtspunkten, um vor allem Rechtssicherheit für die handelnden Personen zu bekommen. Das Feld der Problemstellungen ist extrem breit. Es beginnt bei der Frage, ob Forschungsdaten überhaupt ein urheberrechtlich schützenswertes Gut sind, und reicht über den Datenschutz bis zur Gestaltung der Zusammenarbeit zwischen Institutionen mehrerer Bundesländer bzw. über Ländergrenzen hinweg. Welche Rechte erwirbt zum Beispiel der Betreiber eines Forschungsdatenrepositoriums an den Daten. Wie muss demzufolge die vertragliche Gestaltung zwischen einem Datenlieferanten und einem Betreiber ausgefüllt sein. Für diese und weitere Fragen gibt es bisher kaum Antworten. Sie werden jedoch dringend benötigt und stellen zweifelsfrei einen wesentlichen Teil der Rahmenbedingungen für das Management von Forschungsdaten dar.

### 2.2.3 Absicherung der finanziellen Notwendigkeiten

In einer Phase des Wandels des Wissenschaftsprozesses vom analogen Umgang mit Forschungsdaten zur Erzeugung, Speicherung und der Bereitstellung zur weiteren Nutzung auf digitaler Grundlage bedarf es auch eines Umdenkens bei der finanziellen Absicherung des Prozesses. Es sind teilweise völlig neue Strukturen aufzubauen, ohne die bisherigen Strukturen zu vernachlässigen. Das verlangt zunächst einen deutlich höheren finanziellen Einsatz, der sich erst nach einer Aufbauphase reduzieren wird.

„Aufbauphase: Investitionen für den Aufbau und die Erneuerung geeigneter Forschungsdaten-Infrastrukturen für alle Fachdisziplinen (derzeitiger IT-System- und Daten-Migrationszyklus liegt bei ca. 6 Jahren)

Betriebsphase: Dauerhafte Bereitstellung der Betriebskosten für die Datenbereitstellung und Datenpflege in allen Fachdisziplinen

Angebot flankierender Förderprogramme für Forschungs- und Entwicklungsprojekte beispielsweise für generische Lösungen oder für Standards und Vorgehensmodelle

Bereitstellung von pauschalen Mitteln (Datenpauschale), um die Sicherung der in den jeweiligen Forschungsprojekten entstandenen Daten durch Dritte zu finanzieren und ihre möglichst kostenfreie Zugänglichkeit zu gewährleisten“ (Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz. Kommission Zukunft der Informationsinfrastruktur 2011, S. 52)

Neben der zusätzlichen Bereitstellung von investiven und konsumtiven Mitteln für Betriebskosten und Personal gilt es, die bisherigen Wege der Zusammenarbeit auf nationaler und internationaler Ebene einer Prüfung zu unterziehen. Es

ist zu hinterfragen, ob die gegenwärtigen Geschäftsmodelle, die im Wesentlichen eine kostenfreie Bereitstellung von Serviceleistungen unterstellen, weiterhin tragfähig sind. Die Intensität der Zusammenarbeit der Forschenden auf lokaler, regionaler, nationaler und internationaler Ebene hat in den letzten Jahren, insbesondere durch die digitalen Technologien, eine völlig neue Größenordnung erreicht. Die Frage nach der Übernahme der anfallenden Kosten sowohl für die Infrastruktur als auch für den darauf angebotenen Service ist also durchaus berechtigt. Denkbar wären sehr unterschiedliche Geschäftsmodelle allein für Deutschland, wobei die nachfolgende Aufzählung keinen Anspruch auf Vollständigkeit erhebt:

- Die Kosten werden vollständig von Bund und den Ländern getragen.
- Die Kosten sind vollständig durch die betreibende Einrichtung abzudecken.
- Die Kosten werden durch eine Umlage gemeinschaftlich getragen.
- Die Kosten sind auf die „Verursacher“ (die Forschenden) umzulegen.
- Die Kosten werden durch die Nutzenden übernommen.

Um nicht falsch verstanden zu werden, sei an dieser Stelle deutlich betont, dass der Autor dieses Beitrages sehr wohl für eine kostenlose Bereitstellung von entsprechenden Services plädiert. Dies ist jedoch im Einklang mit der ebenso wichtigen Gewährung der Nachhaltigkeit zu sehen.

## **2.3 Organisatorische Dimension des Forschungsdatenmanagements**

Die organisatorische Dimension zeichnet sich dadurch aus, dass sie das Zusammenwirken der jeweils Beteiligten am FDM betrachtet. Es läuft hinaus auf die grundsätzliche Frage: Welche Strukturen sind zu schaffen und wie kann eine entsprechende Arbeitsteilung zwischen den Infrastruktureinrichtungen zum einen und den Forschenden und dem Servicepersonal zum anderen effizient gestaltet werden?

### **2.3.1 Servicestrukturen**

Die Schaffung von Servicestrukturen stellt sich meist als ein generischer Prozess dar, der in vielen Fällen durch das Engagement der beteiligten Forschenden oder des jeweiligen Servicepersonals geprägt wird. Selten ist ein Top-down-Ansatz zu beobachten, sondern die Entwicklung und der Betrieb von



zum Beispiel Forschungsdatenrepositorien (FDR) sind durch Bottom-up-Prozesse geprägt. Anders als bei der Mehrzahl der Publikationsrepositorien, die meist institutionellen Charakter tragen, handelt es sich bei FDR um Community gesteuerte Aktivitäten. Das hat den Vorteil, dass sie stark durch die praktischen Anforderungen bestimmt sind, aber den Nachteil, dass eine institutionelle Verankerung und damit die Nachhaltigkeit weniger im Fokus stehen.

Der Aufbau von Servicestrukturen ist zwar auf die einzelne Infrastruktureinrichtung bezogen zunächst ein lokaler Prozess, der jedoch jeweils eine internationale, nationale und regionale Betrachtung erfordert, um eine im wissenschaftlichen Gesamtzusammenhang stehende Einordnung zu erfahren. Es würde den Umfang dieses Beitrages sprengen, an dieser Stelle weiter ins Detail zu gehen.

### 2.3.2 Policies zum Forschungsdatenmanagement

Zur organisatorischen Dimension des FDMs sind die Auseinandersetzung und letztlich die Beschlussfassung von Policies für den Umgang mit Forschungsdaten, im Sinne von Leitlinien des Handelns, zu zählen. Die entsprechende Policy der Humboldt-Universität zu Berlin (2014a), die für eine ganze Reihe von Forschungseinrichtungen in Deutschland beispielgebend war, gibt Antworten auf folgende Fragen:

- Wer trägt die Verantwortung für den Umgang mit Forschungsdaten im Rahmen eines Forschungsvorhabens?
- In welcher Form bereitet man Forschungsdaten auf?
- Wer bestimmt den Zeitpunkt und die rechtlichen Rahmenbedingungen für die Veröffentlichung der Forschungsdaten?
- Wo und wie lange sollten die Forschungsdaten eines Forschungsvorhabens aufbewahrt werden?
- Wozu verpflichtet sich die Humboldt-Universität in diesem Zusammenhang?

In diesem Sinne erfüllt eine Policy zum einen das Ziel, den Forschenden der jeweiligen Wissenschaftseinrichtung eine Grundorientierung für den Umgang mit Forschungsdaten zu geben und zum anderen allen Außenstehenden zu vermitteln, dass es innerhalb der Forschungseinrichtung eine bewusste Auseinandersetzung zum Thema Forschungsdaten gegeben hat und weiterhin gibt.

In den vergangenen Jahren ist eine Vielzahl von Schriften sowohl zu Grundaussagen von Policies als auch zu ihrer Gestaltung erschienen. Es sei an dieser Stelle darauf verwiesen und auf eine Diskussion verzichtet (ANDS, o. J.; Rücknagel 2015; Green, Macdonald & Rice 2015).

### 2.3.3 Interne Strategie und Anreizsysteme

Mit einer Policy bekennt sich eine Forschungseinrichtung zu einer Strategie im Umgang mit Forschungsdaten. In vielen Fällen gibt die Policy jedoch nur die Grundzüge für das Management vor, sodass detailliertere Aussagen im Sinne von Handlungsanweisungen benötigt werden (Humboldt-Universität zu Berlin 2014b). Wichtig ist es, an dieser Stelle deutlich zu machen, wo und durch wen die Forschenden Hilfestellungen bei der Bewältigung der vielfältigen Aktivitäten zu erwarten haben. Die HRK-Empfehlungen (Hochschulrektorenkonferenz 2015) beschreiben in diesem Zusammenhang zum Beispiel Szenarien des FDMs und reagieren damit auf den unterschiedlichen Charakter der unterschiedlichen Forschungsprozesse, den Bogen vom Promotionsvorhaben bis zur Forschung in Kooperation mit der Industrie spannend.

Darüber hinaus bedarf es der Festlegung finanzieller Rahmenbedingungen innerhalb der Forschungseinrichtung und einer Prioritätensetzung für mögliche interne Anreizsysteme. Auf diese Weise sollen die Wege zur Umsetzung der Strategie des FDMs sichtbar, die Realisierbarkeit unterstrichen und die Motivation der Forschenden in die gewünschten Bahnen gelenkt werden.

## 2.4 Technische Dimension des Forschungsdatenmanagements

Die technische Dimension des FDMs ist recht vielschichtig und im Rahmen dieses Beitrages nur begrenzt darstellbar. Im Wesentlichen gilt es zu unterscheiden zwischen:

- den Anforderungen an die Forschungsdaten und den Umgang mit ihnen,
- den Anforderungen an die Metadaten über die Forschungsdaten und den sie erzeugenden Forschungsprozessen und
- den Anforderungen an die Aufbewahrung, Speicherung und Bereitstellung der Forschungsdaten (in der Mehrzahl der Fälle in einem Repository).

### 2.4.1 Anforderungen an die Daten und den Umgang mit ihnen

Die nationalen und internationalen Forschungsförderinstitutionen verweisen an dieser Stelle in ihren Papieren sehr gern auf die Einhaltung von Standards und Regelungen der jeweiligen Community, wie zum Beispiel die DFG:

„Wesentlich ist die Einhaltung von Standards. Entsprechende Regelungen sind von den Wissenschaftlern in Kooperation mit Informationsspezialisten zu treffen. Hier sind vor allem auch internationale Regelungen zu berücksichtigen, damit die Kompatibilität auch international gewährleistet ist.“ (Deutsche Forschungsgemeinschaft. Unterausschuss für Informationsmanagement 2009)

Sucht man jedoch nach den Standards und Regelungen, so wird man nur in sehr wenigen Fachgebieten wirklich fündig. In der Mehrzahl der Fälle steht man am Anfang eines Diskussions- und Lernprozesses, sodass zunächst recht allgemein gehaltene Vorgaben die einzige Orientierung darstellen. Lee, Strong, Kahn und Wang (2002) unterscheiden vier Kategorien von Informationsqualität (IQ), in dem sie Intrinsic IQ, Contextual IQ, Representational IQ und Accessibility IQ näher ausführen und dabei auf die Grundlagen einer seriösen wissenschaftlichen Arbeitsweise abheben. Im Mittelpunkt stehen allgemeine Anforderungen wie Exaktheit, Konsistenz, Korrektheit, Glaubwürdigkeit, Verlässlichkeit, Objektivität, Mehrwert, Relevanz, Vollständigkeit, Aktualität, Angemessenheit, Verständlichkeit, Interpretierbarkeit, Prägnanz, Zugänglichkeit, Verfügbarkeit des Systems, Funktionen und Operationen, Bedienbarkeit, System- und Datensicherheit. Ohne Zweifel haben alle diese Anforderungen ihre Berechtigung. Versucht man jedoch, etwas mehr den technischen Aspekt der Diskussion zur Entwicklung von Standards in den Fokus zu rücken, so gewinnt die Nähe der Datenveröffentlichung zu einer klassischen Textpublikation an Relevanz. Folgende Anforderungen sehen dabei Schirnbacher und Müller (2009):

- Zugänglichkeit,
- Nachhaltigkeit,
- Nachvollziehbarkeit,
- Authentizität,
- Qualitätssicherung,
- Bewertung,
- Geschwindigkeit,
- Vollständigkeit.

Eine ausführliche Erläuterung dieser Kriterien und die praxisrelevanten Konsequenzen daraus finden sich in der angegebenen Literaturquelle. Mit diesen Kriterien kommt man allerdings nicht automatisch zu Standards, denn sie stellen nur potenzielle Eckdaten dar. Es bedarf dazu einer detaillierten Auseinandersetzung in der jeweiligen Community, wie man sie zum Beispiel bei Jensen (2012) finden kann.

#### **2.4.2 Anforderungen an die Metadaten über Forschungsdaten und die sie erzeugenden Forschungsprozesse**

Forschungsdaten haben im Wesentlichen die vier Funktionen:

- Nachweisinstrument für die erbrachte Forschung,
- Grundlage für die Kommunikation und Diskussion innerhalb der Forschercommunity,

- Basis für die Anerkennung in der Community (zumindest in ausgewählten Wissenschaftsdisziplinen),
- Nachnutzung durch andere Forschende.<sup>1</sup>

Diese Funktionen, jede für sich mit entsprechender Tragweite, können jedoch nur zur Wirkung kommen, wenn eine Beschreibung der Forschungsdaten adäquat durch Metadaten erfolgt. Je detaillierter und aussagekräftiger dies erfolgt, je mehr werden die Forschungsdaten zu einem anerkannten Gegenstand der Auseinandersetzung innerhalb der Forschercommunity. Es existiert gegenwärtig noch kein allgemeingültiger Metadatenstandard für Forschungsdaten und wahrscheinlich wird es ihn auch künftig nur jeweils bezogen auf das Fachgebiet bzw. die Untersuchungsmethode geben, weil anders als bei einer traditionellen wissenschaftlichen Publikation die Forschungsdaten heterogen und deutlich komplexer sind. Die Ergebnisse der doch recht zahlreichen Befragungen in der letzten Zeit belegen dies sehr deutlich (Umfragen zum Umgang mit Forschungsdaten, o. J.; Kindling, Schirnbacher & Simikovic 2013; Paul-Stüve, Rasch & Lorenz 2015). Das Dublin Core Metadata Set ist für Forschungsdaten nur bedingt geeignet, weil sich insbesondere der Charakter des Entstehungsprozesses der Forschungsdaten nur unzureichend abbilden lässt.

In der einschlägigen Fachliteratur (Jensen, Katsanidou & Zenk-Möltgen 2011; Digital Curation Centre 2004–2017; Farnel & Shiri 2014) unterscheidet man im Allgemeinen vier Hauptgruppen von Metadaten, die hier nur in einem groben Überblick dargestellt werden können:

- Deskriptive Metadaten  
dienen der inhaltlichen Beschreibung der Forschungsdaten und sind somit Voraussetzung für ein thematisch orientiertes Auffinden.
- Strukturelle Metadaten  
beschreiben den Prozess der Gewinnung der Forschungsdaten, um den Nutzenden über das entsprechende Verfahren und seine Randbedingungen zu informieren.
- Administrative Metadaten  
klären die Randbedingungen für den Umgang mit Forschungsdaten. Unter anderem enthalten sie die notwendigen Informationen zum Rechtemanagement und geben Auskunft zur Provenance

---

<sup>1</sup> Es würde innerhalb dieses Artikels zu weit führen, diese Funktionen näher zu erläutern. Sie wurden im Rahmen der Vorlesung „Ausgewählte Aspekte digitaler Informationsversorgung“ am Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin erarbeitet.

- Technische Metadaten  
umfassen alle technischen Parameter der jeweiligen Dateien und sollen so die Gewähr bieten, dass man aufgefundene Forschungsdaten auch weiter verarbeiten kann.

### **2.4.3 Anforderungen an die Aufbewahrung, Speicherung und Bereitstellung der Forschungsdaten**

In der Mehrzahl der Fälle werden Forschungsdaten in einem Repository aufbewahrt. Es unterscheidet sich in erster Linie von einem reinen Dateisystem mit angemessener Ordnerstruktur dadurch, dass der verantwortliche Betreiber den Nutzerinnen und Nutzern, wobei darunter sowohl die die Daten einspeisenden Forschenden als auch die Suchenden zu verstehen sind, spezielle Services zur Verfügung stellt, die einzelne Arbeitsgänge erleichtern sollen. So findet man vielfach Eingabemasken, Verwaltungstools, spezielle Suchmaschinen u. ä.

Es existieren sehr unterschiedliche Repositorien, deren Charakter in der Regel durch den Sammelauftrag bestimmt wird. Pampel, Goebelbecker und Vierkant (2012) unterscheiden fünf Typen:

- Disziplinäre Forschungsdaten-Repositorien,
- Institutionelle Forschungsdaten-Repositorien,
- Projektspezifische Forschungsdaten-Repositorien,
- Multidisziplinäre Forschungsdaten-Repositorien,
- Portale, die verteilte Datensammlungen zugänglich machen.

Bei re3data.org findet sich ein Verzeichnis der weltweit agierenden Forschungsdatenrepositorien. Im Dezember 2016 waren nahezu 1 800 Repositorien gelistet. Die Anforderungen an solche FDR sind äußerst vielschichtig. Re3data gruppiert sie in sieben Aspekte:

- Allgemeine Informationen,
- Services,
- Policies,
- Rechtliche Fragen,
- Technische Standards,
- Metadatenstandards,
- Qualitätsstandards.

Mit dem Data Seal of Approval (o. J.) gibt es eine Leitlinie mit 16 Anforderungen an die Gestaltung von FDR.

### 3 Zusammenfassung

Die drei Dimensionen des FDMs sollen die Diskussion von Themenschwerpunkten erleichtern und klarer als bisher darstellen, welche Zielgruppe sich durch die Fragestellungen angesprochen fühlen sollte und wo die entsprechenden Verantwortlichkeiten liegen. Natürlich wird bei einer Detailbetrachtung deutlich, dass die Grenzen zwischen den Dimensionen keine scharfen Trennlinien verkörpern, sondern schwimmend verlaufen, was insbesondere beim Übergang von der wissenschaftspolitischen Dimension zur organisatorischen Dimension nicht zu übersehen ist.

Das FDM wird gegenwärtig national und international sehr intensiv diskutiert. Es gibt eine entsprechende Arbeitsgruppe innerhalb der Deutschen Initiative für Netzwerkinformation (DINI) für Deutschland, mit ANDS (Australian National Data Service) eine Organisation in Australien, um nur zwei Beispiele zu nennen, und mit der Research Data Alliance ein weltweit agierendes Gremium. Diese Initiativen beschäftigen sich jeweils mit vielen der aufgeführten Fragestellungen und streben nach geeigneten Lösungsmöglichkeiten. Es ist das Ziel des Beitrages, in dieser Auseinandersetzung eine Systematik vorzuschlagen, die eine Weitergabe erreichter Ergebnisse deutlich erleichtern könnte und vor allem eine bessere Vergleichbarkeit gestatten würde.

Kommt man zurück zu den eingangs erwähnten Big Data, so sollte deutlich geworden sein, dass die Gemeinsamkeiten mit Forschungsdaten differenziert zu sehen sind. Innerhalb der technischen Dimension sind die Fragestellungen sehr ähnlich, weil die Anforderungen an die Daten, ihre Metadaten und ihre Aufbewahrung einander gleichen. Deutlich wird dies auch in den teilweise identischen Methoden zur Analyse der Daten, worauf jedoch in diesem Beitrag nicht eingegangen werden konnte.

Nicht ganz so groß sind die Schnittmengen bei der organisatorischen Dimension. FDM erwartet eine spezielle Infrastruktur, die disziplinspezifische Elemente berücksichtigt und insbesondere die Nachnutzung erleichtern soll. Natürlich gibt es ebenso im Bereich von Big Data Datenzentren, bei denen die Aufbewahrung der Daten jedoch wesentlich mehr im Mittelpunkt steht als der Service für die Community. Das ist sicher sehr stark verallgemeinert und hält einem Vergleich im Detail nicht immer stand. Nach dem Verständnis des Autors gibt es die größten Unterschiede bei den Fragestellungen im Rahmen der wissenschaftspolitischen Dimension. Forschungsdaten dienen dem Nachweis der erbrachten wissenschaftlichen Leistung und dienen der Nachnutzung bzw. Fortführung der Forschung, um nur zwei Funktionen von Forschungsdaten an dieser Stelle zu wiederholen. Sie werden als bewahrenswertes Kulturgut verstanden und sollten somit im Fokus des Staates stehen. Big Data können für

sich zumindest in ihrer Breite diesen Anspruch nicht erheben. Sie haben in der Gesellschaft berechtigterweise ihre Anerkennung in vielen Bereichen gefunden und gemeinsam mit neueren Methoden der Data Analytics zu beachtenswerten Ergebnissen geführt. In der Mehrzahl der Fälle stehen jedoch deutlich mehr privatwirtschaftliche Interessen im Vordergrund als gesamtgesellschaftliche. Das soll nicht als eine Geringschätzung von Big Data verstanden werden, sondern lediglich als ein Unterschied zu Forschungsdaten.

Zusammenfassend lässt sich ohne jeden Zweifel feststellen, dass in der Beschäftigung mit digital vorliegenden Daten ein immenses Potenzial steckt, das es auszuschöpfen gilt.

## Literatur

- Australian National Data Service (ANDS). (o.J.). *Data management policy*. Abgerufen von <http://ands.org.au/datamanagement/policy.html>
- Australian Research Council. (2015). *Research data management*. Abgerufen von <http://www.arc.gov.au/research-data-management>
- Big Data. (2013). In *Informatiklexikon*. Abgerufen von <https://www.gi.de/service/informatiklexikon/detailansicht/article/big-data.html>
- Bundeskanzlerin Merkel: Daten sind die Rohstoffe des 21. Jahrhunderts. (2016, 12. März). *CBIT News*. Abgerufen von <http://www.cebit.de/de/news/bundeskanzlerin-merkel-daten-sind-die-rohstoffe-des-21-jahrhunderts.xhtml>
- Data Seal of Approval. (o.J.). [Homepage]. Abgerufen von <http://www.datasealofapproval.org/en/>
- Deutsche Forschungsgemeinschaft. Unterausschuss für Informationsmanagement. (2009, Januar). *Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten*. Abgerufen von [http://www.dfg.de/download/pdf/foerderung/programme/lis/ua\\_inf\\_empfehlungen\\_200901.pdf](http://www.dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901.pdf)
- Digital Curation Centre. (2004–2017). *List of Metadata Standards*. Abgerufen von <http://www.dcc.ac.uk/resources/metadata-standards/list>
- Digitale Weiterbildung: Mittelstand sieht Nachholbedarf bei der Datenanalyse. (2015, 23. November). *Cebra*. Abgerufen von <https://www.cebra.biz/news/praxis/23-11-2015-mittelstand-sieht-nachholbedarf-bei-der-datenanalyse/408488/>
- Farnel, S. & Shiri, A. (2014). Metadata for research data: Current practices and trends. In *Proceedings of the International Conference on Dublin Core and Metadata Applications 2014*. Abgerufen von <http://dcpapers.dublincore.org/pubs/article/viewFile/3714/1937>
- Green, A., Macdonald, S. & Rice, R. (2009, Mai). *Policy-making for research data in repositories: A guide data share* (version 1.2). Abgerufen von <http://www.disc-uk.org/docs/guide.pdf>
- High Level Expert Group on Scientific Data. (2010, Oktober). *Riding the wave: How Europe can gain from the rising tide of scientific data; Final report of the High Level Expert Group on Scientific Data, October 2010*. Abgerufen von [http://ec.europa.eu/information\\_society/newsroom/cf/itemlongdetail.cfm?item\\_id=6204](http://ec.europa.eu/information_society/newsroom/cf/itemlongdetail.cfm?item_id=6204)

- Hochschulrektorenkonferenz. (2014). *Management von Forschungsdaten: Eine zentrale strategische Herausforderung für Hochschulleitungen; Empfehlung der 16. Mitgliederversammlung der HRK am 13. Mai 2014 in Frankfurt am Main*. Abgerufen von [https://www.hrk.de/uploads/tx\\_szconvention/HRK\\_Empfehlung\\_Forschungsdaten\\_13052014\\_01.pdf](https://www.hrk.de/uploads/tx_szconvention/HRK_Empfehlung_Forschungsdaten_13052014_01.pdf)
- Hochschulrektorenkonferenz. (2015). *Wie Hochschulleitungen die Entwicklung des Forschungsdatenmanagements steuern können: Orientierungspfade, Handlungsoptionen, Szenarien; Empfehlung der 19. Mitgliederversammlung der HRK am 10. November 2015 in Kiel*. Abgerufen von [https://www.hrk.de/uploads/tx\\_szconvention/Empfehlung\\_Forschungsdatenmanagement\\_\\_final\\_Stand\\_11.11.2015.pdf](https://www.hrk.de/uploads/tx_szconvention/Empfehlung_Forschungsdatenmanagement__final_Stand_11.11.2015.pdf)
- Humboldt-Universität zu Berlin. (2014a, 8. Juli). *Forschungsdaten-Policy: Grundsätze zum Umgang mit Forschungsdaten an der Humboldt-Universität zu Berlin (Forschungsdaten-Policy) vom 8. Juli 2014*. Abgerufen von <https://www.cms.hu-berlin.de/de/dl/dataman/infos/policy>
- Humboldt-Universität zu Berlin. (2014b, 21. Juli). *Handlungsempfehlungen in Ergänzung zu den Grundsätzen zum Umgang mit Forschungsdaten an der Humboldt-Universität zu Berlin (Version 1.3)*. Abgerufen von <https://www.cms.hu-berlin.de/de/dl/dataman/infos/policy/guidelines/guidelines>
- Jensen, U. (2012). *Leitlinien zum Management von Forschungsdaten: Sozialwissenschaftliche Umfragedaten* (GESIS-Technical Reports, 2012,07). Köln: GESIS – Leibniz-Institut für Sozialwissenschaften. Abgerufen von [http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis\\_reihen/gesis\\_methodenberichte/2012/TechnicalReport\\_2012-07.pdf](http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2012/TechnicalReport_2012-07.pdf)
- Jensen, U., Katsanidou, A. & Zenk-Möltgen, W. (2011). Metadaten und Standards. In S. Büttner, H.-C. Hobohm & L. Müller (Hrsg.), *Handbuch Forschungsdatenmanagement*. Bad Honnef: Bock + Herchen. urn:nbn:de:kobv:525-opus-2412
- Joint Information System Committee. (2016, 11. April). *Managing research data in your institution*. Abgerufen von <https://www.jisc.ac.uk/guides/research-data-management>
- Kindling, M. & Schirmbacher, P. (2013) Die digitale Forschungswelt als Gegenstand der Forschung. *Information, Wissenschaft & Praxis*, 64(2/3), 127–136. Abgerufen von doi:10.1515/iwp-2013-0017
- Kindling, M., Schirmbacher, P. & Simukovic, E. (2013). Forschungsdatenmanagement an Hochschulen: Das Beispiel der Humboldt-Universität zu Berlin. *LIBREAS. Library Ideas*, 23. urn:nbn:de:kobv:11-100212700
- Lee, Y. W., Strong, D. M., Kahn, B. K. & Wang, R. D. (2002). AIMQ: A methodology for information quality assessment. *Information & Management*, 40(2), 133–146. Abgerufen von doi:10.1016/S0378-7206(02)00043-5
- National Science Foundation. (o. J.). *Dissemination and sharing of research results*. Abgerufen von <https://www.nsf.gov/bfa/dias/policy/dmp.jsp>
- Pampel, H., Goebelbecker, H.-J. & Vierkant, P. (2012). re3data.org: Aufbau eines Verzeichnisses von Forschungsdaten-Repositoryn; Ein Werkstattbericht. In B. Mittermaier (Hrsg.), *Vernetztes Wissen: Daten, Menschen, Systeme* (S. 61–73). Jülich: Verlag des Forschungszentrums Jülich. Abgerufen von <http://hdl.handle.net/2128/4699>
- Paul-Stüve, T., Rasch, G. & Lorenz, S. (2015). *Ergebnisse der Umfrage zum Umgang mit digitalen Forschungsdaten der Christian-Albrechts-Universität zu Kiel (2014)*. Zenodo. Abgerufen von <http://www.uni-kiel.de/vfu/de/forschungsdatenmanagement/umfrage-zum-umgang-mit-forschungsdaten>



- Rat für Informationsinfrastrukturen. (2016). *Leistung aus Vielfalt: Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland*. Göttingen: RfII. Abgerufen von <http://www.rfii.de/?wpdmdl=1998>
- re3data.org. (o. J.) *Registry of Research Data Repositories*. [Homepage]. Abgerufen von <http://www.re3data.org>
- Rücknagel, J. (2015). *Policies von Forschungsdatenrepositorien: Eine Analyse*. Nicht veröffentlichte Masterarbeit, Humboldt-Universität zu Berlin.
- Schirmbacher, P. & Müller, U. (2009). *Das wissenschaftliche Publizieren: Stand und Perspektiven*. Abgerufen von <http://edoc.hu-berlin.de/cmsj/32/schirmbacher-peter-7/PDF/schirmbacher.pdf>
- Umfragen zum Umgang mit Forschungsdaten an wissenschaftlichen Institutionen. (o. J.). In *forschungsdaten.org*. Abgerufen von [http://www.forschungsdaten.org/index.php/Umfragen\\_zum\\_Umgang\\_mit\\_Forschungsdaten\\_an\\_wissenschaftlichen\\_Institutionen](http://www.forschungsdaten.org/index.php/Umfragen_zum_Umgang_mit_Forschungsdaten_an_wissenschaftlichen_Institutionen)
- Van der Graaf, M. & Waaijers, L. (2012). *A surfboard for riding the wave: Towards a four country action programme on research data; A knowledge exchange report* (2<sup>nd</sup> ed.). Abgerufen von <http://www.knowledge-exchange.info/surfboard>
- Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz. Kommission Zukunft der Informationsinfrastruktur. (2011, April). *Gesamtkonzept für die Informationsinfrastruktur in Deutschland: Empfehlungen der Kommission Zukunft der Informationsinfrastruktur im Auftrag der Gemeinsamen Wissenschaftskonferenz des Bundes und der Länder*. Berlin: Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz. Abgerufen von [http://www.leibniz-gemeinschaft.de/fileadmin/user\\_upload/downloads/Infrastruktur/KII\\_Gesamtkonzept.pdf](http://www.leibniz-gemeinschaft.de/fileadmin/user_upload/downloads/Infrastruktur/KII_Gesamtkonzept.pdf)
- Wissenschaftsrat. (2012). *Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020* (Drs. 2359–12). Berlin. Abgerufen von <http://www.wissenschaftsrat.de/download/archiv/2359-12.pdf>

Alle Internetquellen wurden zuletzt am 07.01.2017 aufgerufen.