

Dienstleistungen für den Aufbau und die Verzeichnung von Forschungsdatenrepositorien - Anforderungen aus der Praxis des KIT -

Bachelorarbeit im Studiengang Bibliotheks- und Informationsmanagement
an der Hochschule der Medien Stuttgart

vorgelegt von

Rouven Schabinger

am

14.03.2014

Erstprüfer: Prof. Magnus Pfeffer

Zweitprüfer: Dr. Hans-Jürgen Goebelbecker

Eidesstattliche Versicherung

Hiermit versichere ich, Rouven Schabinger, an Eides Statt, dass ich die vorliegende Bachelorarbeit mit dem Titel: „Dienstleistungen für den Aufbau und die Verzeichnung von Forschungsdatenrepositorien - Anforderungen aus der Praxis des KIT -“ selbstständig und ohne fremde Hilfe verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen wurden, sind in jedem Fall unter Angabe der Quelle kenntlich gemacht. Die Arbeit ist noch nicht veröffentlicht oder in anderer Form als Prüfungsleistung vorgelegt worden. Ich habe die Bedeutung der eidesstattlichen Versicherung und die prüfungsrechtlichen Folgen (§ 26 Abs. 2 Bachelor-SPO (6 Semester), § 23 Abs. 2 Bachelor-SPO (7 Semester) bzw. § 19 Abs. 2 Master-SPO der HdM) sowie die strafrechtlichen Folgen (gem. § 156 StGB) einer unrichtigen oder unvollständigen eidesstattlichen Versicherung zur Kenntnis genommen.

Auszug aus dem Strafgesetzbuch (StGB)

§ 156 StGB Falsche Versicherung an Eides Statt

Wer von einer zur Abnahme einer Versicherung an Eides Statt zuständigen Behörde eine solche Versicherung falsch abgibt oder unter Berufung auf eine solche Versicherung falsch aussagt, wird mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft.

.....

Ort, Datum

.....

Unterschrift

Kurzfassung

Forschungsdatenrepositorien nehmen im Management von Forschungsdaten eine zentrale Rolle ein. Durch diese wird eine publikationsunabhängige Nutzung von Forschungsdaten möglich und gleichzeitig werden Aspekte der Langzeitarchivierung berührt. Für deren Aufbau und den Nachweis werden Dienstleistungen benötigt, die lokale Einrichtungen des Informationsmanagements (z.B. die Bibliothek) für ihre Forscher anbieten können.

Aus diesem Grund wurden am KIT 2012 und 2013 exemplarisch mehrere Erhebungen durchgeführt, um den Bedarf zu ermitteln und Schritte für die Optimierung des Managements von Forschungsdatenrepositorien einzuleiten. Seit 2012 wird an der KIT-Bibliothek in Kooperation mit anderen Forschungseinrichtungen das Such- und Nachweisinstrument für Forschungsdatenrepositorien re3data.org entwickelt. Durch die Ansiedlung der Entwicklergruppe von re3data.org an der KIT-Bibliothek besteht für Forschungsgruppen am KIT die besondere Möglichkeit, an der aktuellen Ausgestaltung von re3data.org zu partizipieren: Durch persönliche Kontakte entstehen Mehrwerte sowohl bei den Forschungsgruppen, als auch bei der re3data.org-Entwicklergruppe selbst. Diese können für den Aufbau von Forschungsdatenrepositorien und für die Optimierung von re3data.org genutzt werden.

Die Arbeit bietet einen Überblick über den Status Quo des Forschungsdatenmanagements am KIT.

Schlagwörter:

Forschungsdatenrepositorien, Forschungsdatenmanagement, KIT, re3data.org, Erhebung

Abstract

Research data repositories play a crucial role in the management of research data. They are a source of readily accessible research data independent from publications while also affecting aspects of long-term curation. Organizing and collecting those repositories require services that can be provided by local departments of information management of a research institution such as a library.

On this account several surveys were exemplarily conducted at KIT in 2012 and 2013 to evaluate the necessity and to take actions for optimizing the management of research data repositories. The tool re3data.org for the collection and description of research data repositories is being developed by KIT cooperatively with other research institutions since 2012. By locating the development team of re3data.org at KIT library, the research departments of KIT are able to closely participate in the ongoing design process of re3data.org. Surplus values are generated through face-to-face contact for both the research departments as well as for the re3data.org developers themselves. Such values can be utilized for setting up research data repositories and for the optimization of re3data.org.

This thesis provides an overview of the actual approaches in research data management at KIT.

Keywords:

Research data repositories, research data management, KIT, re3data.org, study

Inhaltsverzeichnis

Eidesstattliche Versicherung.....	2
Kurzfassung.....	3
Abstract.....	4
Inhaltsverzeichnis.....	5
Abbildungsverzeichnis.....	7
Abkürzungsverzeichnis.....	8
Danksagung.....	10
1 Einleitung.....	11
2 Theorie.....	13
2.1 Forschungsdaten.....	13
2.2 Forschungsdatenmanagement.....	14
2.2.1 Lebenszyklus.....	15
2.2.2 Metadaten / Standards.....	19
2.2.3 Policies / Förderorganisationen.....	20
2.2.4 Recht.....	23
2.2.5 Publikation / Referenzierung.....	24
2.2.6 Langzeitarchivierung.....	26
2.3 Forschungsdatenrepositorien.....	27
2.3.1 Definition.....	27
2.3.2 Klassifizierung.....	31
2.3.3 Zertifizierung.....	32
2.3.4 Dienstleistungen.....	34
3 Praxis.....	36
3.1 KIT.....	36
3.2 re3data.org.....	38
3.3 Vergleich mit Großbritannien.....	40
4 Erhebungen.....	43
4.1 Befragung 2012.....	43
4.2 Befragung 2013.....	47

Inhaltsverzeichnis	6
4.2.1 Methodik.....	47
4.2.2 Umfrageergebnisse.....	48
4.2.3 Interview 1.....	52
4.2.4 Interview 2.....	54
4.2.5 Interview 3.....	56
4.2.6 Interview 4.....	57
5 Analyse.....	60
6 Ausblick.....	66
Quellenverzeichnis.....	67

Abbildungsverzeichnis

Abb. 1: Die Anzahl der Datensätze des „Big Data“ ist klein, die meisten Forschungsdaten sind „Small Data“.....	14
Abb. 2: DCC Curation Lifecycle Model.....	17
Abb. 3: Domains, data stores and curation boundaries.....	18
Abb. 4: Schichten und Komponenten einer DMS-Architektur.....	30
Abb. 5: Aspects of a Research Data Repository with the corresponding icons used in re3data.org.....	39
Abb. 6: Bibliothekarische Aufgaben in den Projekten des Jisc-RDM-Programms 2011–13.....	40
Abb. 7 (links): Datentypen und Datenformate.....	43
Abb. 8 (rechts): Gesamtmenge der gespeicherten Daten in einer OE.....	43
Abb. 9: Erwartete Probleme bei der LZA.....	46

Abkürzungsverzeichnis

API	Application Programming Interface
DCC	Digital Curation Centre
DFG	Deutsche Forschungsgemeinschaft
DINI	Deutsche Initiative für Netzwerkinformation
DMP	Data Management Plan(-ning)
DOI	Digital Object Identifier
DSA	Data Seal of Approval
FDR	Forschungsdatenrepositorium engl. RDR: research data repository
FDM	Forschungsdatenmanagement engl. RDM: research data management
HGF	Helmholtz-Gemeinschaft
ICSU WDC	International Council for Science - World Data Center
KIT	Karlsruher Institut für Technologie engl. Karlsruhe Institute of Technology
LSDF	Large Scale Data Facility
LZA	Langzeitarchivierung, meist als digitale LZA (dLZA)
nestor	Network of Expertise in long-term Storage and availability of digital Resources in Germany
OAIS	Open Archival Information System

OE	Organisationseinheit (dazu zählen Zentren, Fakultäten bzw. Lehrstühle und Institute am KIT)
PI	Persistent Identifier
re3data.org	Registry of Research Data Repositories
SCC	Steinbuch Centre for Computing am KIT
ZAK-CODIGT	Zentrum für Angewandte Kulturwissenschaft - Center of Digital Tradition am KIT

Danksagung

Ich möchte mich bei meinen Betreuern Herrn Pfeffer und Herrn Goebelbecker bedanken, unter deren Anleitung die Arbeit entstanden ist. Ein weiterer Dank geht an Herrn Ralf Schneider vom ZAK-CODIGT und an die Abteilung Forschungsdienste der KIT-Bibliothek, vor allem an Frau Claudia Kramer. Es freut mich auch, dass einige Forscher an der Umfrage teilgenommen haben und bereit waren, ihre Zeit für Interviews zu verwenden.

Ohne die Unterstützung meiner Eltern, meines Bruders und meiner Großeltern wäre die Arbeit nicht möglich gewesen; besonders mein Opa hatte immer ein offenes Ohr für mich. Ich möchte mich auch bei meinen Freunden Juri, Marco, Kai und Thomas für das geduldige Zuhören bedanken.

1 Einleitung

*"Was? du suchst? du möchtest dich verzehnfachen, verhundertfachen? du suchst Anhänger?
- Suche Nullen [und Einsen]!"*

Friedrich Nietzsche: Sprüche und Pfeile, 14 / Götzen-Dämmerung

[Ergänzung des Verfassers]

Die Zukunft der Wissenschaft hängt maßgeblich auch von der Nutzung von Informationstechnologien ab, die heute schon im Alltag der Forscher wachsende Bedeutung erfahren. Die so genannte E-Science bietet den Forschern nicht nur bei wissenschaftlichen Experimenten bzw. Simulationen und der kollaborativen Vernetzung Unterstützung, sondern hat auch Auswirkung auf die Speicherung und Veröffentlichung der Daten im Sinne einer data driven science. Dass diese Forschungsdaten ein Management benötigen, um ihr volles Potential zu entfalten, liegt auf der Hand. Damit ergeben sich gerade für Infrastruktureinrichtungen wie Bibliotheken oder Rechenzentren neue Aufgaben, die es zu lösen gilt.¹

Ein wichtiger Teil dieses Themas ist die Konkretisierung der Bemühungen um Forschungsdatenmanagement (FDM) in so genannten Forschungsdatenrepositorien (FDRs). Durch diese wird eine Nachnutzung der Daten unabhängig von wissenschaftlichen Textpublikationen möglich. Gleichzeitig stehen Forschungsdatenrepositorien auf der Schwelle zum Themenbereich der digitalen Langzeitarchivierung (LZA), die teils gleiche, teils andere Anforderungen und Akteure hat und in dieser Arbeit eine nebengeordnete Rolle spielt.

Zielsetzung dieser Arbeit ist es herauszufinden, welche Bedürfnisse Wissenschaftler am KIT (Karlsruher Institut für Technologie) bzgl. des Aufbaus

¹ Siehe Blanke u.a. (2013): Bericht über das Symposium zum zehnjährigen Bestehen der Abteilung Forschung & Entwicklung, IV., S. 50; Vgl. Schulze; Stockmann (2013): Open Science und Networked Science, S. 32

und des Nachweises von FDRs haben. Weiter sollen Handlungsansätze erarbeitet werden, auf welche Weise wissenschaftliche Dienstleistungseinrichtungen wie die KIT-Bibliothek dem Anforderungsprofil gerecht werden können. Zu diesem Zweck wurde am KIT bereits 2012 eine Umfrage zu Forschungsdaten und LZA unter einigen Instituten durchgeführt. Diese Umfrage wurde 2013 fortgesetzt und auf Repositorien spezifiziert. Zusätzlich hat der Verfasser mit vier Beteiligten vertiefende Interviews geführt.

Die Arbeit gliedert sich in vier Teile: einen theoretischen, einen praktischen, einen empirischen und einen analytischen Teil mit Empfehlungscharakter.

Nach diesem Kapitel 1 (Einleitung) führt das Kapitel 2 (Theorie) in das Thema Forschungsdaten ein, gibt grundlegende Definitionen an und reißt die Aspekte des FDM an, welche für das Betreiben und den Nachweis von Repositorien besonders relevant sind. Zusätzlich werden die Themen FDRs und Dienstleistungen im FDM dargestellt. Weiter werden im Kapitel 3 (Praxis) das KIT, das FDR-Such- und Nachweisinstrument re3data.org, sowie internationale Praktiken des FDM am Beispiel Großbritanniens vorgestellt. Das dann folgende Kapitel 4 (Erhebungen) ist der Darstellung der Ergebnisse aus den Umfragen und den Interviews gewidmet. Auf deren Basis und den Erkenntnissen aus den vorausgegangenen Kapiteln wird in Kapitel 5 (Analyse) eine Analyse durchgeführt, die Folgerungen und Empfehlungen für das weitere Vorgehen am KIT zum Ziel hat. Die Arbeit schließt mit einem Ausblick auf verwandte Themen und zukünftige Entwicklungen.

In dieser Arbeit wird nur die männliche Geschlechtsbezeichnung verwendet. Sofern nicht ausdrücklich darauf hingewiesen wird, sind dadurch beide Geschlechter gemeint.

2 Theorie

2.1 Forschungsdaten

„Die verschiedenen Wissenschaften sprechen von Messdaten, Rohdaten, empirischen Daten, Quelldaten, Forschungsrohdaten etc., verwenden viele unterschiedliche Begriffe und haben auch unterschiedliche Konzepte von dem, was für sie ein „Datum“ ist. Mittlerweile kristallisiert sich heraus, dass im deutschen Sprachgebrauch der Begriff „Forschungsdaten“ bevorzugt wird.“²

Etwas seltener werden die Begriffe Primärdaten und Originaldaten gebraucht; im Englischen sind die Termini research data oder scientific data gebräuchlich. Eine universelle Definition von Forschungsdaten ist aufgrund der Vielfalt der Anwendungsbereiche schwierig und kann daher nur grob erfolgen. Man kann versuchen sich der Thematik über die Forschungsmethodik zu nähern:

„Unter Forschungsdaten sind [...] digitale und elektronisch speicherbare Daten zu verstehen, die im Zuge eines wissenschaftlichen Vorhabens z.B. durch Quellenforschungen, Experimente, Messungen, Erhebungen oder Befragungen entstehen.“³

oder über die Art der Daten:

„[...] “research data” are defined as factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings. A research data set constitutes a systematic, partial representation of the subject being investigated.“⁴

2 Büttner u.a. (2011): Research Data Management, S. 15

3 Winkler-Nees (2010): Forschungsdaten besser nutzen: S. 2

4 OECD (2007): Principles and Guidelines for Access to Research Data from Public

Im Allgemeinen liegt der Fokus auf digitalen Daten (digital born data), die oft mit dem Term „digitale Objekte“ beschrieben werden. Ausgehend von einer physischen über eine logische bis hin zu einer konzeptuellen Ebene können das dann einzelne Dateien oder Dateiformate (Text, Multimedia), komplexere Einheiten wie Internetseiten und Datenbanken oder auch Software und ganze Betriebssysteme sein.⁵

Die Datenvolumina können je nach Forschungsprojekt variieren, es sollte aber sensibel mit dem Begriff „Big Data“ umgegangen werden, da Forschungsdaten in vielen Fällen nur einige Gigabyte groß sind (siehe Abb. 1).⁶

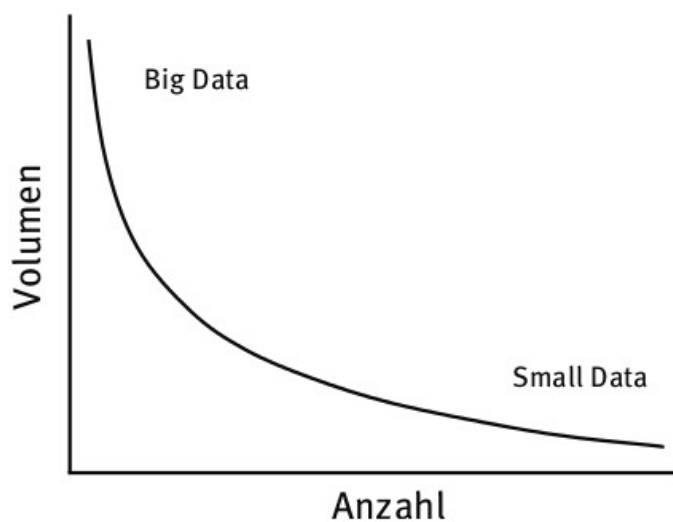


Abb. 1: Die Anzahl der Datensätze des „Big Data“ ist klein, die meisten Forschungsdaten sind „Small Data“⁷

2.2 Forschungsdatenmanagement

Wie in der Einleitung angedeutet, benötigen Forschungsdaten in einer datenzentrierten Wissenschaft ein systematisches Management. Damit sind sowohl Infrastrukturen und standardisierte Vorgehensweisen als organisa-

Funding, S. 13

5 Siehe Funk (2010): Digitale Objekte und Formate, Kap. 7:3

6 Siehe Klump; Bertelmann (2013): D 8 Forschungsdaten, S. 574

7 Grafik aus: ebenda, S. 574

torische Gegebenheiten und sich ergebende Dienstleistungen gemeint. Die Verarbeitung und Speicherung heterogener Datenmengen unter der Beteiligung vieler unterschiedlicher Wissenschaftsdisziplinen ist als Prozess zu begreifen, der hohe Anforderungen stellt. Damit ergeben sich nicht nur technische Fragestellungen, sondern es werden auch weitergehende Aspekte im Sinne einer wissenschaftlichen Integrität adressiert. Diese Aufgabe muss im Rahmen einer Forschungsdateninfrastruktur national und lokal gelöst werden.⁸

Bereiche des Forschungsdatenmanagements (FDM, engl. research data management RDM), die einen gesonderten theoretischen Hintergrund erfordern, sind nach der Einführung des Lebenszyklus als eigene Unterkapitel aufgeführt.

2.2.1 Lebenszyklus

Um ein nachhaltiges Datenmanagement von Forschungsdaten zu gewährleisten, erscheint es sinnvoll, die kurativen Maßnahmen am Forschungsprozess und damit am Lebenszyklus (engl. lifecycle) der Daten auszurichten.⁹ Mit digital curation können sowohl Tätigkeiten des FDM als auch der LZA gemeint sein. Bestehende Probleme mit dem gegenwärtigen Publikationsprozess von Forschungsdaten haben dazu geführt, dass mehrere Lebenszyklusmodelle mit unterschiedlichen Ansätzen entwickelt wurden.¹⁰ Als prominentester Vertreter wird das Curation Lifecycle Model des Digital Curation Centre (DCC) vorgestellt, welches das ideale Verhalten beim Datenmanagement fokussiert und einen Überblick über das Data Curation Continuum des Australian National Data Service gegeben, welches vornehmlich die existierenden Zusammenhänge im Zuge der Veröffentlichung beschreibt.

8 Siehe Büttner u.a. (2011): Research Data Management, S. 14

9 Siehe Higgins (2012): The Lifecycle of Data Management, S. 17 f.

10 Vgl. Rümpel (2011): Der Lebenszyklus von Forschungsdaten, S. 25 ff.

Das DCC Lifecycle Model (siehe Abb. 2) unterscheidet sequentielle Maßnahmen (z.B. ingest), dargestellt durch den äußeren Ring, andauernde Maßnahmen (z.B. preservation planning), dargestellt durch die vier inneren Ringe und gelegentliche Maßnahmen (z.B. dispose), dargestellt durch die Pfeile außerhalb des Diagramms.¹¹ In der Conceptualize-Phase sollten das Management, der im geplanten Forschungsvorhaben entstehenden Daten in Bezug auf ihre Erhebungsart, Beschreibung und Identifizierung sowie auf ihre Speicherung festgelegt werden (z.B. durch einen Data Management Plan - DMP). Schon früh können so Probleme mit syntaktischer und semantischer Interoperabilität umgangen werden, wenn Standardisierung berücksichtigt wird. Die direkte Einbindung eines Information Professionals ermöglicht Ratschläge zu Best Practices und Hinweise auf geltende Standards (z.B. auch für die Ablage in einem Repository).¹² Bei der Create-or-Receive-Phase werden die Rohdaten erhoben und ggf. mit technischen (oder anderen) Metadaten versehen. Der Information-Manager sollte auf die Beachtung von Policies hinweisen.¹³ Im weiteren Schritt werden die Daten in Bezug auf ihre Speicherdauer (nur während des Projekts oder LZA) selektiert (appraise and select) oder ganz verworfen (dispose). So können die kurativen Maßnahmen auf die notwendigen Daten fokussiert werden, auch wenn die Speicherung aller Daten möglich wäre.¹⁴ Mit dem Ingest wird die Übergabe an ein Repository, Archiv oder eine Ablage im weiteren Sinn bezeichnet und die Verbindung zum OAIS-Modell (Open Archival Information System) geschaffen. Hier kommen Sicherheitsprüfungen, Qualitätskontrollen und rechtliche Bestimmungen zum Tragen.¹⁵

11 Siehe Digital Curation Centre (o. J.): DCC Curation Lifecycle Model

12 Siehe Higgins (2012): The Lifecycle of Data Management, S. 23 f.

13 Siehe ebenda, S. 25 ff.

14 Siehe ebenda, S. 28 f.

15 Siehe ebenda, S. 30 f.

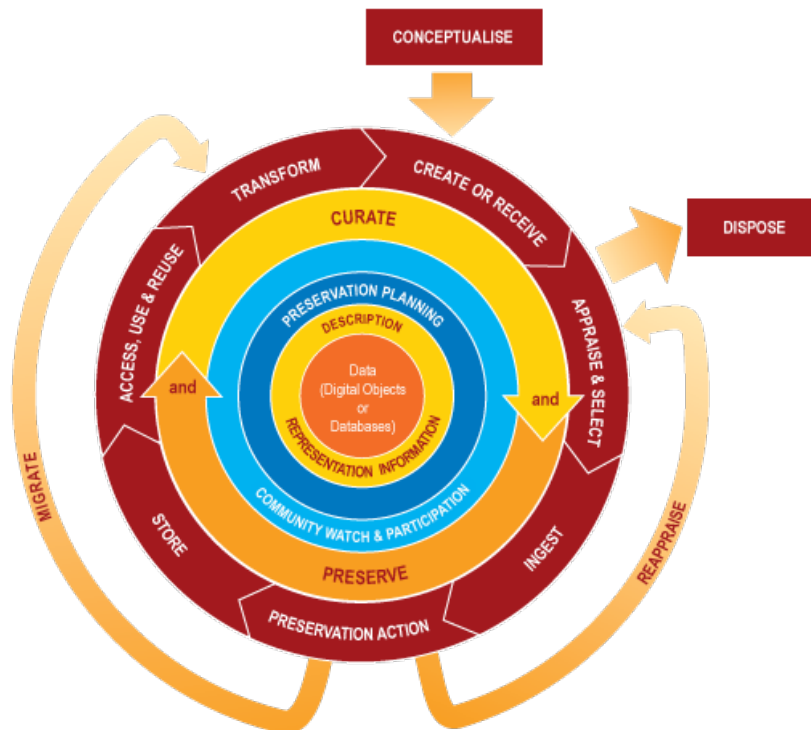


Abb. 2: DCC Curation Lifecycle Model¹⁶

Es folgen preservation action und store wie z.B. persistente Adressierung, Metadatenanreicherung und Konsistenzprüfungen.¹⁷ Bei access, use and reuse wird die anvisierte Zielgruppe für die Daten, nötige Authentifizierungs- und Autorisierungsmechanismen, Lizenzen und Modalitäten der Datenabfrage relevant. Weiter kann es im Sinne einer Nachnutzung zur Weiterverarbeitung der bestehenden Daten kommen (transform), welche die Grundlage für weitere Forschung bietet und den Zyklus schließt. Es wird empfohlen, ggf. die Rohdaten und die bearbeiteten Daten vorzuhalten.¹⁸

Neben den genannten sequentiellen und gelegentlichen Tätigkeiten, gibt es Maßnahmen, die über den ganzen Lebenszyklus hin durchgeführt wer-

¹⁶ Grafik aus: Digital Curation Centre (o. J.): DCC Curation Lifecycle Model

¹⁷ Siehe Higgins (2012): The Lifecycle of Data Management, S. 32

¹⁸ Siehe ebenda, S. 37 f.

den sollten: description and representation Information, preservation planning, community watch and participation, curate and preserve.¹⁹

Das Data Curation Continuum des Australian National Data²⁰ unterscheidet dagegen verschiedene Domains (siehe Abb. 3):

- Private research domain: „There may be a large number of data objects which are updated frequently. At this stage, researchers typically manage their own data. Preservation and metadata may not be needed, and access to the data is limited.“
- Public domain: „There are likely to be a smaller number of selected static data objects which have accrued more metadata, and which may be managed and preserved through institutional arrangements such as repositories. This data is more likely to be publicly accessible, possibly in association with print publications.“

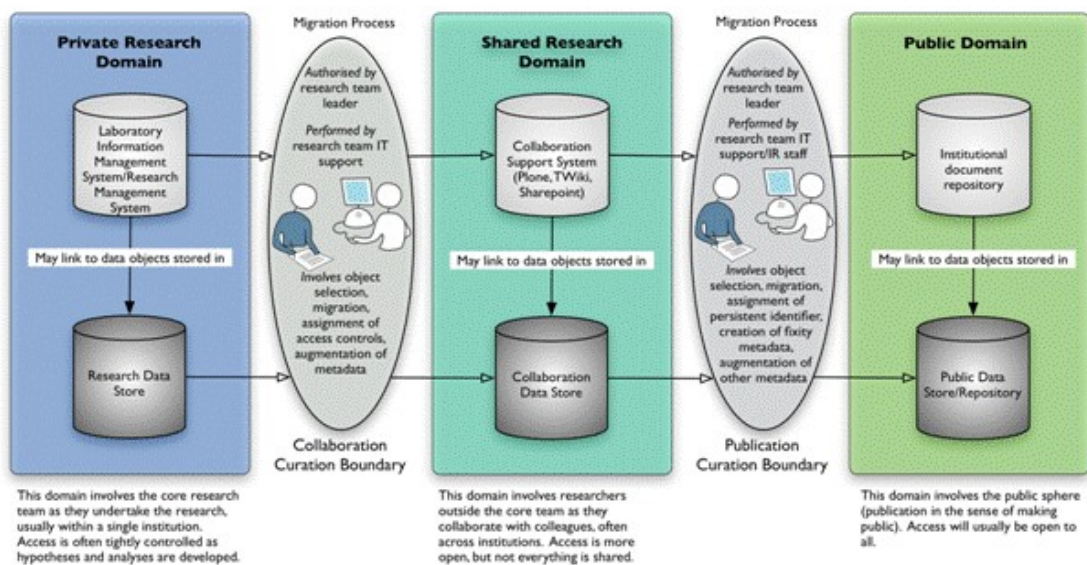


Abb. 3: Domains, data stores and curation boundaries²¹

Zwischen den beiden liegt die shared research domain, bei der die Daten beim Ein- und Austritt eine Migration / Transformation erfahren. Hier wer-

¹⁹ Digital Curation Centre (o. J.): DCC Curation Lifecycle Model

²⁰ Siehe Australian National Data Service (o. J.): Curation Continuum

²¹ Grafik aus: ebenda

den kurative Maßnahmen und Verantwortlichkeiten der eingebundenen Information Professionals wichtig und die Zielgruppe (z.B. Forscher aus der gleichen Fachrichtung) muss festgelegt werden.²²

2.2.2 Metadaten / Standards

„Metadaten sind strukturierte Daten zur einheitlichen Beschreibung von Ressourcen jeglicher Art (z.B. Daten, Dokumente, Personen, Gemälde, Orte, Gebäude, Konzepte). Sie sorgen dafür, dass den Beschreibungen von Instanzen ein- und derselben Ressourcenklasse eine einheitliche Struktur zugrunde liegt und erleichtern so das Suchen, Finden und Selektieren relevanter Ressourcen aus einer Vielzahl möglicher Ressourcen.“²³

Es gibt verschiedene Arten von Metadaten: z.B. identifizierende / beschreibende, administrative oder technische,²⁴ deren Erfassung manuell, semi- oder vollautomatisch (z.B. bei technischen Metadaten) erfolgen kann. Aus der Definition lässt sich ableiten, dass Metadaten eine sinnvolle Nachnutzung von Forschungsdaten erst ermöglichen. Damit fungieren sie als Bindeglied zwischen den Forschungsdaten und deren Infrastruktur (im engeren Sinne Repositorien) und beeinflussen damit die Qualität der Forschungsdaten.²⁵ Analog zu Forschungsdaten existiert auch im Bereich der Metadaten eine Vielfalt,²⁶ die Standardisierung notwendig macht, um nicht zuletzt für die digitale LZA (dLZA) eine Interoperabilität und Kompatibilität

22 Siehe ebenda

23 Rühle; Kompetenzzentrum Interoperable Metadaten (o. J.): Kleines Handbuch Metadaten, S. 2

24 Siehe ebenda, S. 2

25 Siehe Kindling (2013): Qualitätssicherung im Umgang mit digitalen Forschungsdaten, S. 144

26 Vgl. Riley (2010): Seeing Standards

zu gewährleisten.²⁷ Damit geht die Tendenz einher, offene Standards (z.B. XML) für die Metadatenformate zu verwenden.²⁸

Als grundlegende Metadatenstandards können der Dublin Core²⁹ oder das DataCite Metadata-Schema³⁰ verwendet werden, die für die Beschreibung von Forschungsdaten mindestens vorhanden sein sollten. Daneben können fachspezifische Metadaten-Frameworks verwendet werden: z.B. in den Sozialwissenschaften DDI (Data Documentation Initiative) oder in der Klimaforschung CERA-2 (Climate and Environmental Retrieval and Archive).³¹ Zur weiteren inhaltlichen Erschließung können auch im bibliothekarischen Bereich etablierte Klassifikationen (z.B. DDC – Dewey-Decimal-Classification) oder mehrsprachige Thesauri zum Tragen kommen.³²

Wichtige Metadatenstandards in der digitalen LZA sind PREMIS (Preservation Metadata: Implementation Strategies), LMER (Long-term preservation Metadata for Electronic Resources) und METS (Metadata Encoding and Transmission Standard),³³ welche im Falle von METS auch direkt in das OAIS-Referenzmodell eingebunden werden können (siehe Kapitel 2.2.6).³⁴

2.2.3 Policies / Förderorganisationen

Die am FDM beteiligten Akteure (Förderorganisationen, Fachgesellschaften, Infrastruktureinrichtungen, Verlage etc.) haben begonnen Verhaltensgrundsätze in so genannten Data Policies (Verpflichtungserklärungen) zu

27 Siehe Jensen u.a. (2011): Metadaten und Standards, S. 91

28 Vgl. Hein (2012): Metadaten für die Langzeitarchivierung, S. 106

29 Siehe Deutsche Forschungsgemeinschaft (2009): Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten, S. 3

30 Vgl. Digital Curation Centre (o. J.): General Research Data

31 Siehe Jensen u.a. (2011): Metadaten und Standards, S. 86 f.

32 Siehe ebenda, S. 88

33 Siehe Schumann (2012): Einführung in die digitale Langzeitarchivierung, S. 44

34 Vgl. Menne-Haritz (2005): METS, Abschnitt: Einleitung

verbriefen, die sich in interdisziplinäre, disziplinäre, institutionelle und Journal- bzw. Editorial Policies gliedern lassen.³⁵

Die DFG (Deutsche Forschungsgemeinschaft) verlangt disziplinübergreifend: „Primärdaten als Grundlagen für Veröffentlichungen sollen auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, zehn Jahre lang aufbewahrt werden.“³⁶ Weiter fordert sie Antragssteller direkt auf, das Datenmanagement für ihr Forschungsvorhaben zu beachten und fördert durch ihr Programm „Wissenschaftliche Literaturversorgungs- und Informationssysteme“ (LIS) gezielt den Aufbau überregionaler Informationsinfrastrukturen,³⁷ also auch von FDRs. Die Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen (Allianzinitiative) rät zur Verwendung von disziplinären Standards (z.B. für Metadaten), die auch eine interdisziplinäre Nutzung ermöglichen.³⁸ Die DINI (Deutsche Initiative für Netzwerkinformation) betont eine enge disziplinorientierte Kooperation auch zwischen Infrastruktureinrichtungen (Bibliotheken, Daten- und Rechenzentren).³⁹

Auf disziplinärer Ebene lassen sich die Bermuda Principles anführen, die eine schnelle Veröffentlichung der Daten, auch vor der Publikation betonen.⁴⁰ Natürlich müssen solche disziplinären und interdisziplinären Policies auch auf institutioneller Ebene umgesetzt werden, um ihre volle Wirkung zu entfalten.

35 Siehe Pampel; Bertelmann (2011): „Data Policies“ im Spannungsfeld zwischen Empfehlung und Verpflichtung, S. 49

36 Deutsche Forschungsgemeinschaft (2013): Vorschläge zur Sicherung guter wissenschaftlicher Praxis, S. 21

37 Siehe Deutsche Forschungsgemeinschaft (o. J.): Sicherung von Forschungsdaten

38 Siehe Schwerpunktinitiative „Digitale Information“ der Allianz der Wissenschaftsorganisationen (2010): Grundsätze zum Umgang mit Forschungsdaten

39 Siehe Deutsche Initiative für Netzwerkinformation (2009): Positionspapier Forschungsdaten, S. 7

40 Siehe Wellcome Trust (o. J.): Statement on genome data release

Oft wird auch der Open Access mit den Forschungsdaten verknüpft: „Open access contributions include original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material“,⁴¹ auch auf europäischer⁴² und internationaler⁴³ Ebene. Auch die Helmholtz-Gemeinschaft (HGF) schließt Forschungsdaten in ihre Open-Access-Bewegung ein und bietet bspw. Webinare dazu an.⁴⁴

Wie bei der Textpublikation ist auch bei zusätzlicher Veröffentlichung der Daten eine inhaltliche Qualitätssicherung in Form eines Peer Review nötig, was dazu führt, dass auch Verlage sich mit Policies auseinandersetzen.⁴⁵ Auch haben Zeitschriften mit einer expliziten Richtlinie zum Umgang mit Daten einen höheren Impact-Faktor.⁴⁶ Die Journal-Policies fokussieren meist die Ablage in einem geeignetem Repository oder eine persistente Referenzierung unter Berücksichtigung des Fachbereichs.⁴⁷ Es kann auch Open Access als Voraussetzung zur Veröffentlichung der Daten bei dem Journal gefordert werden.⁴⁸

41 Max-Planck-Gesellschaft (2003): Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities

42 Siehe EUROHORC; ESF (2008): The EUROHORCs and ESF Vision on a Globally Competitive ERA and their Road Map for Actions to Help Build It, S. 4

43 Siehe OECD (2007): OECD Principles and Guidelines for Access to Research Data from Public Funding, S. 15

44 Siehe Helmholtz-Gemeinschaft (2013): Open Access

45 Siehe Pampel; Bertelmann (2011): „Data Policies“ im Spannungsfeld zwischen Empfehlung und Verpflichtung, S. 55

46 Siehe Weichselgartner u.a. (2011): Archivierung von Forschungsdaten, S. 192; nach Piwowar; Chapman (2008): A review of journal policies for sharing research data

47 Vgl. American Geophysical Union (2013): AGU Publications Data Policy; Cell (o. J.): Information for Authors, Abschnitt: Distribution of Materials and Data

48 Siehe PLOS One (o. J.): PLOS Editorial and Publishing, Abschnitt: Methods, and Data

2.2.4 Recht

Im Zuge der Veröffentlichung und Nachnutzung von Forschungsdaten werden auch Fragen des Urheber- bzw. Leistungsschutzrechts und des Datenschutzes relevant. Einzelne Forschungsdaten sind in der Regel nicht urheberrechtlich geschützt, da es an der geistigen Schöpfungshöhe für ein Werk fehlt. Auch für Datensammlungen im Sinne von Tabellen oder Datenbanken werden eher logische statt individuelle Gesichtspunkte verwendet. Weiter ist der Datenbankhersteller selten der Träger des wirtschaftlichen und finanziellen Risikos (im Sinne der Vorschrift), sondern die Trägerinstitution, der dann die Leistungsschutzrechte zufielen.⁴⁹ Bei Drittmittelgebern können Verträge, die den Forschern (oder deren Einrichtung) einfache Nutzungsrechte zugestehen, sinnvoll sein.⁵⁰ Bei der Veröffentlichung und damit der Nachnutzung durch Dritte sollte aufgrund des geringen Schutzes der Daten „ein Lizenzvertrag geschlossen werden, der die zulässige Nutzung der Daten durch den Fremdforscher regelt und eine Verschwiegenheitsklausel beinhaltet, die die Fremdforscher verpflichtet, die Daten nicht unbefugt weiterzugeben“.⁵¹ In einem regulären Arbeitsverhältnis werden die Nutzungsrechte, die in der Ausübung der Stelle entstehen, auf den Arbeitgeber übertragen. Bei Professoren ist das nicht immer der Fall, darum sollten auch dort Verträge über die Einräumung von Nutzungsrechten vom Forscher an die Einrichtung geschlossen bzw. in den Arbeitsvertrag integriert werden.⁵²

Bei der Erhebung personenbezogener Daten greifen die Datenschutzgesetze⁵³ und setzen in der Regel eine Einwilligung voraus. Weiter muss ge-

49 Siehe Spindler; Hillegeist (2011): Rechtliche Probleme der elektronischen Langzeitarchivierung von Forschungsdaten, S. 63 f.; Vgl. §3, §4, § 87a-e UrhG

50 Siehe ebenda, S. 65

51 Siehe Spindler; Hillegeist (2010): Langzeitarchivierung wissenschaftlicher Primärdaten, Kap. 16:19

52 Siehe Spindler; Hillegeist (2011): Rechtliche Probleme der elektronischen Langzeitarchivierung von Forschungsdaten, S. 65 f.

53 BDSG und entsprechende Landesgesetze, SGB X

prüft werden, ob alle personenidentifizierenden Merkmale notwendig sind und ggf. eine Anonymisierung erfolgen muss.⁵⁴ Dadurch ergibt sich z.B. in den Sozialwissenschaften im Hinblick auf den Datenzugang ein Spannungsverhältnis zwischen der wissenschaftlichen Auswertbarkeit und dem Gehalt auf der einen Seite und den entsprechenden datenschutzrechtlichen Bestimmungen auf der anderen Seite.⁵⁵ Im Bereich der LZA können sich die rechtlichen Bestimmungen in Bezug auf Vervielfältigungen oder bei medizinischen Daten noch zuspitzen.⁵⁶

2.2.5 Publikation / Referenzierung

In den Naturwissenschaften kann die Veröffentlichung von Forschungsdaten in drei Modelle⁵⁷ eingeteilt werden, welche einem dynamischen Prozess unterliegen und ineinander übergehen können:

- als eigenständiges Objekt in einem Repository (siehe Kapitel 2.3)
- mit einer textuellen Dokumentation
- als Anreicherung einer interpretativen Text-Publikation.

Traditionell werden Forschungsdaten in eine Textpublikation in Form von Abbildungen und Tabellen integriert und in Auszügen im Anhang einer wissenschaftlichen Arbeit angeführt. In Folge des digitalen Fortschritts hat es sich zusätzlich etabliert, bspw. auf Verlagsplattformen die Forschungsdaten mit oder ohne persistenter Adressierung anzubieten. Weiter ist es möglich, die Forschungsdaten zeitgleich zur Verlagsveröffentlichung in ei-

54 Siehe Spindler; Hillegeist (2011): Rechtliche Probleme der elektronischen Langzeitarchivierung von Forschungsdaten, S. 67 f.

55 Siehe Huschka (2011): Datenmanagement und Data Sharing, S. 40

56 Siehe Spindler; Hillegeist (2010): Langzeitarchivierung wissenschaftlicher Primärdaten, S. Kap. 16:4 und 16:8 ff.

57 Siehe Dallmeier-Tiessen (2011): Strategien bei der Veröffentlichung von Forschungsdaten, S. 160

nem externen Repository abulegen.⁵⁸ Hierfür sind Kooperationen zwischen Verlagen und Repositorienbetreibern notwendig, die teilw. schon existieren: Elseviers ScienceDirekt Artikel werden z.B. mit Daten in PANGAEA angereichert,⁵⁹ Dryad arbeitet mit mehreren Verlagen zusammen und hat eine Joint Data Archiving Policy entwickelt (siehe Kap. 2.2.3).⁶⁰

Mit Zunahme der Menge digitaler Objekte wird eine dauerhafte, eindeutige Adressierung und Versionierung notwendig, die über standortgebundene Verweise (z.B. URLs) hinausgeht und standortunabhängige Identifizierungsmechanismen gewährleistet, eine exakte Zitierung der Daten ermöglicht und auch für die LZA wichtig ist. Über so genanntes Resolving ist es dann durch Umleitung der Anfrage möglich, weiterhin auf eine Ressource zuzugreifen, auch wenn sich der physikalische Speicherort verändert.⁶¹

Mit Handle, DOI, ARC (Archival Resource Key), [PURL (Persistent Uniform Resource Locator)] und URN (Uniform Resource Name) bestehen verschiedene Persistent Identifier-Systeme (PI), die sich im Hinblick auf die Art der digitalen Objekte, Land der Registrierungsagentur und Kosten unterscheiden.⁶² ⁶³ Als für Forschungsdaten gut geeignetes Beispiel wird das DOI-System und seine beteiligten Einrichtungen kurz vorgestellt.

Das Digital Object Identifier-System (DOI) wird von der International DOI Foundation betrieben und kooperiert mit Registrierungsagenturen in verschiedenen Ländern und für verschiedene Fachbereiche. Es baut auf dem Handle-System auf und ist ein ISO-Standard.⁶⁴

Bsp: <http://dx.doi.org/10.5524/100001>

58 Siehe ebenda, S. 163 f.

59 Siehe Reed Elsevier (2010): Elsevier and PANGAEA take next step

60 Siehe Dryad (2013): Joint Data Archiving Policy (JDAP)

61 Siehe Von der Hude (2012): Persistent Identifier, S. 129

62 Siehe ebenda, S. 130

63 Vgl. hierzu auch auch das European Persistent Identifier Consortium (EPIC)

64 Siehe International DOI Foundation (o. J.): DOI Handbook, Abschnitt 1.2

Die TIB Hannover ist Mitglied bei der datenzentrierten Registrierungsagentur DataCite und ist Ansprechpartner für deutsche Datenzentren mit naturwissenschaftlich-technischem Schwerpunkt.⁶⁵

2.2.6 Langzeitarchivierung

Unter Langzeitarchivierung (LZA, engl. preservation) – meist verstanden als digitale Langzeitarchivierung (dLZA) – versteht man die über die dauerhafte Speicherung hinausgehende Erhaltung digitaler Ressourcen, um eine zukünftige Nachnutzung und Interpretierbarkeit zu gewährleisten.⁶⁶

„'Langzeit' ist die Umschreibung eines nicht näher fixierten Zeitraumes, währenddessen wesentliche, nicht vorhersehbare technologische und soziokulturelle Veränderungen eintreten [...], die sowohl die Gestalt als auch die Nutzungssituation digitaler Ressourcen in rasanten Entwicklungszyklen vollständig umwälzen können.“⁶⁷

Grundlegend ist dabei die Bindung digitaler Daten an einen Datenträger und die Vergänglichkeit der Datenträgerformate (Bsp. von der Floppy Disk zum USB-Stick). Der Aufwand für die Erstellung digitaler Daten ist gering, die Sicherung der Interpretierbarkeit und Verfügbarkeit aber höher.⁶⁸

„Für ein digitales Archiv spielen die Frage der Authentizität und Integrität von Objekten eine zentrale Rolle“, um die Echtheit der Daten auf Dauer zu gewährleisten.⁶⁹ Oft wird das unter den Begriff der Vertrauenswürdigkeit gefasst. Als theoretisches Referenzmodell für ein Archiv hat sich in der LZA das Open Archival Information System (OAIS) etabliert, welches von amerikanischen Raumfahrtorganisationen Ende der 1990er Jahre entwi-

65 Siehe DataCite (o. J.): German National Library of Science and Technology

66 Siehe Liegmann; Neuroth (2010): Einführung, Kap. 1:3

67 Ebenda, S. Kap. 1:2

68 Siehe Schumann (2012): Einführung in die digitale Langzeitarchivierung, S. 39 f.

69 Siehe ebenda, S. 41 f.

ckelt wurde und die theoretische Modellierung mit Anwendungsfällen verknüpft.⁷⁰

In der LZA können verschiedene Erhaltungsstrategien zum Einsatz kommen:

- Die physische Datensicherung nennt sich Bitstream Preservation.
- Unter Migration / Konversion versteht man die regelmäßige Konvertierung der Objekte in aktuelle Datenformate, z.B. DOC in PDF (anderes Datenformat) oder DOC in DOCX (neuere Version). Vorteile sind Lesbarkeit, Nachteile eventuelle Datenverluste.
- Bei der Emulation wird nicht das digitale Objekt, sondern „die alte Software-Umgebung auf neuer Hardware und in neuer Systemumgebung imitiert (=emuliert).“ Vorteile liegen im Erhalt der Originaldaten, Nachteile im Programmieraufwand.

Die Digitalisierung selbst wird im Allgemeinen nicht als Erhaltungsstrategie angesehen, da auch digitale Objekte einer Erhaltung bedürfen. Wichtig bei der Wahl der Erhaltungsstrategie ist, welche so genannten signifikanten Eigenschaften der digitalen Objekte erhalten bleiben sollen.⁷¹

2.3 Forschungsdatenrepositorien

2.3.1 Definition

Ein Forschungsdatenrepositorium (FDR) bietet die Möglichkeit, Forschungsdaten zeitlich und örtlich unabhängig von einer Textpublikation als eigenständiges Objekt zu veröffentlichen. In manchen Disziplinen ist auch eine automatisierte Publikation aus einem anderen System heraus möglich.⁷² Repositorien spielen eine wichtige Rolle im FDM: Sie können für das

70 Siehe Brübach (2010): Das Referenzmodell OAIS, Kap. 4:3 f.

71 Siehe Schumann (2012): Einführung in die digitale Langzeitarchivierung, S. 42 f.

72 Siehe Dallmeier-Tiessen (2011): Strategien bei der Veröffentlichung von Forschungs-

data sharing oder allgemeiner den gemeinschaftlichen Austausch innerhalb einer wissenschaftlichen Community benutzt werden. Zusätzlich können sie für die LZA von Forschungsdaten verantwortlich sein.⁷³ Die Anzahl der vorhandenen FDRs im Web ist schwer zu quantifizieren, wird aber auf Tausende geschätzt – Tendenz steigend.⁷⁴ Am häufigsten sind FDRs der Fachbereiche Geowissenschaften, Medizin, Biologie, Astronomie und den Sozialwissenschaften zu finden.⁷⁵ Vermutlich finden sich auch viele FDRs in der shared domain (Kap 2.2.1 Abb. 3), die sich aber der öffentlichen Sichtbarkeit entziehen.

Terminologisch werden zahlreiche Begriffe unter einem FDR subsumiert:

„Such infrastructures can include so called data archives, data centers, digital libraries, digital collections and the like. They are collectively summarized under the term Research Data Repositories (RDR).“⁷⁶

Weiter gebräuchlich sind: „Science Data Repositories (SDRs) science data center(s), [science] data repository, scientific data repository, and science digital repository“⁷⁷ und ihre deutschen Äquivalente.

Eine eindeutige Definition, was ein FDR ist, bleibt schwierig, da damit viele unterschiedliche Systeme gemeint sein können und eine gewisse Unschärfe vorherrscht, welche der Dynamik der Thematik Rechnung trägt. Eine mögliche Herangehensweise ist ein Repositorium idealerweise als ein digitales Langzeitarchiv anzusehen. Dieser Anspruch wird zwar nicht von allen FDRs erfüllt, aber er gibt einen Rahmen vor und zeigt, dass Sys-

daten, S. 161

73 Siehe Aschenbrenner; Neuroth (2011): Forschungsdaten-Repositorien, S. 101

74 Siehe Marcial; Hemminger (2010): Scientific data repositories on the Web, S. 2030

75 Siehe Schaaf (2011): Forschungsdaten im Netz, S. 51; Marcial; Hemminger (2010): Scientific data repositories on the Web, S. 2035 Fig 1

76 Pampel u.a. (2013): Making Research Data Repositories Visible, S. 2

77 Marcial; Hemminger (2010): Scientific data repositories on the Web, S. 2030

teme eine technische Komponente wie auch organisatorische Komponenten beinhalten.⁷⁸

„Gemäß OAIS wird unter einem digitalen Langzeitarchiv eine Organisation (bestehend aus Personen und technischen Systemen) verstanden, die die Verantwortung für den Langzeiterhalt und die Langzeitverfügbarkeit digitaler Objekte sowie für ihre Interpretierbarkeit zum Zwecke der Nutzung durch eine bestimmte Zielgruppe übernommen hat.“⁷⁹

Die technischen Funktionen lassen sich auf Speicherkonzepte von digitalen Objekten, Metadatenverwaltung (auch PIs), Datentransfer (Ingest-Workflow, Schnittstellen zur Nachnutzung) und Authentifizierungs- und Autorisierungsmechanismen exemplarisch spezifizieren.⁸⁰ Dabei lässt sich die Systemarchitektur eines solchen Datamanagementsystems (DMS) in Schichten darstellen (siehe Abb. 4). In der Persistenzschicht lassen sich datensatzorientierte (Datenbanken) und dateiorientierte Konzepte unterscheiden.⁸¹ Es können aber auch Hybridsysteme vorkommen sowie Grid- und Cloud-Technologien angewendet werden.⁸² Softwaresysteme für ein FDR sind z.B. EPrints, DSpace, Fedora und iRODS. Zumeist sind das Open-Source-Systeme, die teilw. entwicklungsgeschichtlich aus Dokumentenservern entstanden sind.⁸³

78 Vgl. Aschenbrenner; Neuroth (2011): Forschungsdaten-Repositoryn, S. 101

79 Dobratz; Schoger (2010): Grundkonzepte der Vertrauenswürdigkeit und Sicherheit, Kap. 5:3

80 Siehe Aschenbrenner; Neuroth (2011): Forschungsdaten-Repositoryn, S. 103

81 Siehe Razum (2011): Systeme und Systemarchitekturen für das Datenmanagement, S. 131

82 Siehe ebenda, S. 135

83 Siehe Aschenbrenner; Neuroth (2011): Forschungsdaten-Repositoryn, S. 104 f.

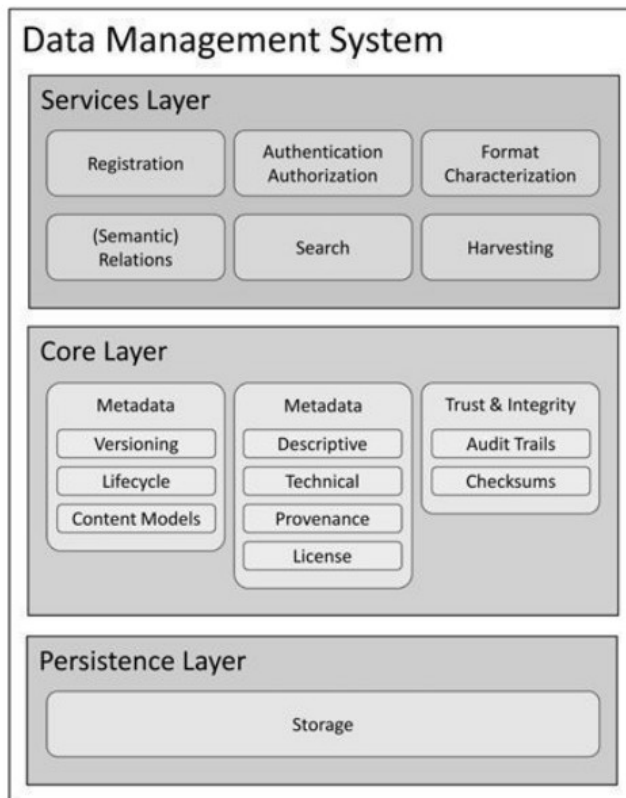


Abb. 4: Schichten und Komponenten einer DMS-Architektur⁸⁴

Trotz der erwähnten Mannigfaltigkeit von FDRs möchte der Verfasser einige Grundanforderungen festhalten und gewisse Systeme ausschließen. Unter der Bedingung, dass FDRs Forschungsdaten (siehe 2.1) enthalten, entfallen so genannte Forschungsdatenbanken „[...]“, die Eckwerte über Forschungsprojekte oder -gruppen, wie zum Beispiel Partner, den Forschungsgegenstand oder finanzielle Aspekte nachweisen⁸⁵ und „verwandte [...] Systeme [...] wie Code-Repositories (vgl. Apache Subversion, Git), Registries (vgl. oft Datenbank-basierte Kataloge wie Service Registries, Metadaten-Registries) und Andere [...]“⁸⁶. Eine weitere Abgrenzung kann zu reinen Dokumentenservern (z.B. OPUS), Nachweiskatalogen und -por-

⁸⁴ Grafik aus: Razum (2011): Systeme und Systemarchitekturen für das Datenmanagement, S. 113 Abb. 2

⁸⁵ Schaaf (2011): Forschungsdaten im Netz, S. 30

⁸⁶ Siehe Aschenbrenner; Neuroth (2011): Forschungsdaten-Repositoryen, S. 102

talen sowie kollaborativen Tools (z.B. CMS, Wikis) gezogen werden, wobei gerade für letztere ein fließender Übergang besteht. Die FDRs sollten online verfügbar sein, Open Access ist optional.

2.3.2 Klassifizierung

In Anlehnung an re3data.org erfolgt eine Klassifizierung von FDRs in Bezug zu dem dort entwickelten und angewandten Schema.⁸⁷

- disziplinäres Repositorium:

„This is a collection of research outputs with a common link to a particular subject discipline. Subject repositories are likely to cover one broad-based discipline, with contributors from many different institutions supported by a variety of funders; the repositories themselves are likely to be funded from one or more sources within the subject community.“⁸⁸

Es gibt neben disziplinären (z.B. PANGAEA) auch FDRs mit interdisziplinärem Anspruch (z.B. figshare).

- institutionelles Repositorium:

„This is a collection of research outputs with a common link to a particular institution, usually by authorship. These repositories are likely to cover more than one research discipline, to have funders in many if not all the Research Councils and support communities who have different approaches to research dissemination. Whether deposit of content is mandatory is a decision that will be made by each institution. The institutions may have many requirements for the content of the repository, from open access dissemination, through metrics,

87 Siehe Vierkant u.a. (2013): Schema for the description of research data repositories, S. 17 f.

88 Jones u.a. (2008): Report of the Subject and Institutional Repositories Interactions study, S. 5

marketing to strategic planning. It is likely that many of these processes in the past were undertaken through collection of bibliographic information."⁸⁹

Mit institutionell kann sowohl eine ganze Einrichtung (bspw. eine Universität, z.B. DSpace@MIT), als auch ein Teil einer Einrichtung (bspw. ein Institut) gemeint sein (z.B. Hamburger Zentrum für Sprachkorpora Korpus Repositorium). Im ersten Fall ist die Ausrichtung des FDR naturgemäß eher interdisziplinär, im zweiten Fall eher disziplinär, wobei auch fließender Übergang zu projektbezogenen FDRs (folgend) bestehen kann.

- andere: (bspw. projektbezogen)

„This is a collection of research outputs with a common link to one (or more) funder(s). These are likely to cover the funder’s remit, which is usually subject-based but can become indistinct at subject boundaries, and will have authors from many institutions. Deposit of content is usually mandatory and can include project-related material, such as completion project reports. The funders will have requirements for the content of repository, from metrics through to strategic planning.”⁹⁰

Ein Beispiel hierfür ist das KCDC (KASCADE Cosmic Ray Data Centre) am KIT (siehe Kapitel 4.2.3)

2.3.3 Zertifizierung

Es kann in einem abstrakteren Schritt eine Übertragung der funktionalen Anforderungen an ein FDR im Sinne einer Qualitätssicherung erfolgen. Es können zwei Ebenen von Qualitätssicherung festgehalten werden: Die Ebene der Forschungsdaten und der verbundenen Forschungsmethodik und - auf aggregierter Ebene - die ihrer Infrastruktur und der damit verbun-

89 Ebenda, S. 5

90 Ebenda, S. 5

denen Dienstleistungen, welche sich gegenseitig beeinflussen.⁹¹ Diese Arbeit wird sich mit der zweiten Ebene beschäftigen und sich auf FDRs fokussieren.

Als Pendant zum DINI-Zertifikat für Dokumenten- und Publikationsservices haben sich im Bereich der Forschungsdaten verschiedene Zertifizierungssysteme etabliert, die auf Vertrauenswürdigkeit als Maßstab für ein Archiv abzielen:⁹²

- Data Seal of Approval (DSA)⁹³
- nestor-Siegel (Network of Expertise in long-term Storage and availability of digital Resources in Germany) für vertrauenswürdige digitale Langzeitarchive (basierend auf DIN 31644)⁹⁴
- Trustworthy Repositories Audit & Certification (TRAC)⁹⁵ des Center for Research Libraries / ISO 16363⁹⁶, der auf dem OAIS-Modell (ISO 14721:2012) basiert.
- Certification of ICSU WDC (International Council for Science - World Data Center) members⁹⁷

Die ersten drei Standards werden als European Framework for Audit and Certification of Digital Repositories in drei Stufen zusammengeführt:

- „Basic Certification is granted to repositories which obtain DSA certification;

91 Siehe Kindling (2013): Qualitätssicherung im Umgang mit digitalen Forschungsdaten, S. 138 und 141, nach Kindling & Schirnbacher, 2013 in diesem Heft S. 134

92 Vgl. Mokrane (2013): Repository Audit and Certification Interest Group DSA, S. 1

93 <http://datasealofapproval.org/en/information/guidelines/>

94 http://www.langzeitarchivierung.de/Subsites/nestor/DE/nestor-Siegel/siegel_node.html?jsessionid=737699C73597AFEB66D5285C97194DCB.prod-worker2

95 <http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/trac>

96 <http://www.iso16363.org/>

97 <http://icsu-wds.org/our-members/membership-application/criteria-membership-certification>

- Extended Certification is granted to Basic Certification repositories which in addition perform a structured, externally reviewed and publicly available self-audit based on ISO 16363 or DIN 31644;
- Formal Certification is granted to repositories which in addition to Basic Certification obtain full external audit and certification based on ISO 16363 or equivalent DIN 31644⁹⁸

Die Research Data Alliance versucht in ihrer Repository Audit and Certification Interest Group langfristig die unabhängig voneinander entwickelten Standards DSA (eher Geisteswissenschaften) und WDS (eher Naturwissenschaften) auf einer Grundstufe zusammenzubringen und zu einem globalen Standard auszubauen.⁹⁹

Weiter existiert eine Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) am DCC.¹⁰⁰

2.3.4 Dienstleistungen

Es wurde ein Pyramiden-Modell mit vier Ebenen von FDM-Aktivitäten von Bibliotheken vorgeschlagen, welches sowohl institutionelle, als auch übergreifende Aufgaben benennt.¹⁰¹ Bspw. werden zu den wichtigeren Aufgaben die Policyentwicklung (national, institutionell) und „develop local data curation capacity“ gezählt¹⁰² - was als FDR-Management und LZA interpretiert

98 Trusted Digital Repository (2010): European Framework for Audit and Certification of Digital Repositories

99 Siehe Mokrane (2013): Repository Audit and Certification Interest Group DSA, S. 2

100 Eine Auflistung der Standards (Repository Audit and Certification Efforts) unter <https://docs.google.com/a/umich.edu/spreadsheet/ccc?key=0Ap-J0U-5ZnEtdGZ-CekJ3Q1Z5Y2I2NnFWemxMQS1NOWc#gid=0>

101 Siehe Corall (2012): Roles and responsibilities S. 110–11; nach Lewis (2010): Libraries and the management of research data, S. 154

102 Auf der von Corall hinzugefügten untersten Ebene wird die Thematik noch einmal durch „develop and manage access to data collections“ adressiert.

tiert werden kann. Weiter spielt eine allgemeine Forschungsdatenberatung, das Vermitteln von research-based learning unter Studierenden und Öffentlichkeitsarbeit bei den Forschern eine Rolle.

Die Dienstleistungen (services), die in den aufgeführten Themenfeldern von FDM (und im Hinblick auf ein FDR) angeboten werden können, lassen sich grob in zwei Dimensionen einteilen: eine zeitliche (strategisch-operativ) und eine funktionale (technisch-organisatorisch). Es bestehen jedoch bei den zeitlichen Aspekten starke Übergänge: Jede strategische Dienstleistung braucht eine operative Umsetzung (z.B. Policy), jede technische Dienstleistung hat eine organisatorische Komponente (z.B. DOI-Vergabe). Weiter gibt es eine Schnittmenge zwischen allen Dienstleistungen, die man als allgemeine Beratung ansehen kann und über die Thematik oder die Existenz der Dienstleistung aufklärt. Im Allgemeinen werden Dienstleistungen für jeden Konsument individuell erbracht, es besteht jedoch eine Verbindung zwischen strategischen Dienstleistungen und dem Nutzen für viele, z.B. re3data.org.¹⁰³

103 Vgl. Umlauf; Lohnert (2011): Lexikon der Bibliotheks- und Informationswissenschaft, S. 91 und 93, Stichwort: „Dienstleistung“

3 Praxis

3.1 KIT

Das Karlsruher Institut für Technologie (KIT) als Zusammenschluss des Forschungszentrums Karlsruhe und der Universität Karlsruhe deckt hauptsächlich naturwissenschaftlich-technische Fächer ab und erzeugt mit seinen über 100 Instituten (ca. 8.000 Mitarbeiter, derzeit ca. 25.000 Studenten) eine Vielzahl von Forschungsdaten. Konkret ist das KIT als Mitglied in der HGF in deren Programmen Energie, Erde und Umwelt, Schlüsseltechnologien und Struktur der Materie aktiv.

Am KIT wurde eine Policy verabschiedet, die sich an den Vorschlägen zur Sicherung guter wissenschaftlicher Praxis der DFG und der HGF orientieren und in Abschnitt 2 e) „Aufbewahrung von Originaldaten“ Folgendes beinhaltet:

„Die oder der für ein Forschungsprojekt Verantwortliche hat sicherzustellen, dass Originaldaten als Grundlagen für Veröffentlichungen, Patente und/oder FuE-Arbeiten auf haltbaren und gesicherten Datenträgern mindestens fünf Jahre nach Abschluss der Vorhaben/Projekte aufbewahrt werden und über die Organisationseinheit für Kontrollorgane zugreifbar sind. Es wird sichergestellt, dass die Kernaussagen einer wissenschaftlichen Arbeit auch über die fünf Jahre hinaus durch Originaldaten belegbar und nachvollziehbar bleiben. Weitergehende Aufbewahrungspflichten aufgrund gesetzlicher Bestimmungen sowie Maßnahmen zum Schutz personenbezogener Daten bleiben hiervon unberührt.“¹⁰⁴

Interessant ist auch die Entwicklung der Forschungsdokumentation am KIT in einer historischen Perspektive. Von 1978 bis 1993 wurden am Cam-

¹⁰⁴ Karlsruher Institut für Technologie (2010): Regeln zur Sicherung guter wissenschaftlicher Praxis im Karlsruher Institut für Technologie (KIT), S. 5

pus Nord bzw. dessen Vorgängereinrichtung (Kern-)forschungszentrum Karlsruhe über 10.000 so genannte Primärberichte erstellt. Diese „formalisierte Zwischenform zwischen den direkt gewonnenen Forschungsdaten und einer nachfolgenden Publikation“ birgt auch heute noch relevante, datenintensive Experimentbeschreibungen z.B. im Bereich Energieforschung. Im Zuge eines umfassenden FDM hat die KIT-Bibliothek bereits ein digitales Verzeichnis dieser Berichte erstellt und möchte mittelfristig alle Primärberichte digital in einem Repository zur Verfügung stellen.¹⁰⁵

Die KIT-Bibliothek betreut mit der Abteilung Forschungsdienste seit 2010 campusübergreifend die Themen Forschungsdaten (bzw. FDM), Digitalisierung und LZA.¹⁰⁶ Der operative Schwerpunkt liegt derzeit auf FDM – in den Bereichen Digitalisierung und LZA kooperiert die KIT-Bibliothek nur punktuell mit anderen zentralen Einrichtungen des Informationsmanagements, die unter verschiedenen Blickwinkeln entsprechende Fragestellungen bearbeiten:

- Das Rechenzentrum / Steinbuch Centre for Computing (SCC) betreibt mit der Large Scale Data Facility (LSDF) intensive Speicherung von Forschungsdaten und Datenlebenszyklusanalyse. Neben Bitstreaming werden Forschungs- und Serviceansätze zur dLZA verfolgt.¹⁰⁷
- Das am Zentrum für Angewandte Kulturwissenschaft und Studium Generale (ZAK) angesiedelten Center of Digital Tradition (CODIGT) widmet sich der Aufgabe der dLZA zur Erhaltung des digitalen Kulturerbes sowie für Daten aus Wissenschaft und Forschung in Form komplexer digitaler Objekte.¹⁰⁸

105 Siehe Goebelbecker (2010): „Nie war er so wertvoll wie heute.“

106 Siehe KIT-Bibliothek (2014): Forschungsdatenmanagement am KIT

107 Siehe Steinbuch Centre for Computing (2013): Large-Scale Data Management & Analysis

108 Siehe Zentrum für Angewandte Kulturwissenschaft (2013): Center of Digital Tradition

- Das KIT-Archiv stellt das Langzeitgedächtnis des KIT im Sinne eines klassischen Archivs dar und kooperiert mit dem ZAK-CODIGT bzgl. dLZA.¹⁰⁹

Es bestehen weiter auch Kooperationsprojekte mit dem auf dem Campus Nord gelegenen FIZ (Leibniz-Institut für Informationsinfrastruktur) im Themenbereich Informationsmanagement. Eine spezielle Zusammenarbeit auf dem Gebiet des FDM wird von beiden Seiten angestrebt.

Die Bibliothek betreibt zwei zentrale Dokumentenserver (EVA STAR, KAROLA), die demnächst zu einem System zusammengeführt werden und dem Nachweis von wissenschaftliche Publikationen dienen.

3.2 re3data.org

Um bestehende FDRs nachzuweisen, versucht die KIT-Bibliothek durch das zusammen mit der Humboldt-Universität Berlin und dem GeoForschungszentrum Potsdam geschaffene DFG-Projekt zum Aufbau eines "Registry of Research Data Repositories" (re3data.org)¹¹⁰ ein weltweites, interdisziplinäres, webbasiertes Verzeichnis zu etablieren.¹¹¹ Zwar gibt es ähnliche Initiativen, die aber andere Schwerpunkte haben: Das Datenbank-Infosystem (DBIS) der Universitätsbibliothek Regensburg verzeichnet hauptsächlich bibliografische Datenbanken und nur marginal FDRs. The Directory of Open Access Repositories – OpenDOAR versucht Ähnliches, begrenzt sich aber auf bei der Zugänglichkeit auf Open Access. Für BioSharing, ZENODO von OpenAIRE und DataCite bestehen Kooperatio-

109 Siehe KIT-Archiv (2012): Startseite

110 <http://www.re3data.org/>

111 Siehe Pampel u.a. (2013): Making Research Data Repositories Visible: The re3data.org Registry, S. 7

nen mit re3data.org;¹¹² für DataBib ist eine intensive Kollaboration ange-
dacht.

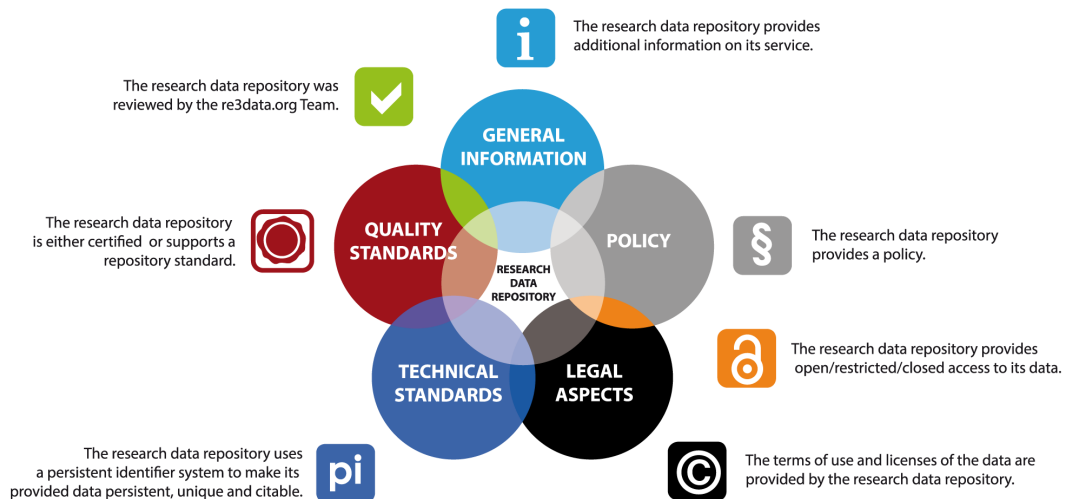


Abb. 5: Aspects of a Research Data Repository with the corresponding icons used in re3data.org¹¹³

Derzeit sind in re3data.org 650 FDRs verzeichnet, von den knapp 600 im eigenen Schema erschlossen wurden. - Tendenz steigend. Dieses Schema bildet die Eigenschaften eines FDR möglichst umfänglich und signifikant ab. Die wichtigsten Kriterien werden anhand verschiedener Icons visualisiert (siehe Abb. 5): Allgemeine Informationen (z.B. Kurzbeschreibung, Inhaltstypen und Schlagwörter), Verantwortlichkeiten, Policies, rechtliche Aspekte (z.B. Lizenzen), technische Standards / Metadaten-Standards (z.B. APIs, FDR-Software, PIs), Qualitätsstandards (z.B. Zertifikate).¹¹⁴

¹¹² re3data.org team (2013): BioSharing and re3data.org cooperate on the collection and description of data repositories; re3data.org team (2013): Memorandum of Understanding between OpenAIRE; re3data.org team (2012): Memorandum of Understanding between DataCite and re3data.org

¹¹³ Grafik aus: Pampel u.a. (2013): Making Research Data Repositories Visible, S. 2 Fig. 1

¹¹⁴ Siehe ebenda, S. 7 f.

3.3 Vergleich mit Großbritannien

In Großbritannien steht mit dem Joint Information Systems Committee (Jisc) eine zentrale Fördereinrichtung für Informations- und Kommunikationstechnologie bereit,¹¹⁵ die auch das DCC initiiert hat.¹¹⁶ So unterstützt es in Kooperation mit dem DCC bspw. in einem vom 2011 bis 2013 laufenden Programm Universitäten, FDM einzuführen oder zu verbessern.¹¹⁷

	Projekt-admin.	Bedarfs-analyse	Policy	DMP	Meta-daten	Rep-Man	Training	Service
data.bris	-	-	x	-	x	x	x	x
RoaD-MAP	x	x	x	-	-	-	x	x
ADMIRe	-	-	-	-	x	x	x	x
Orbital	-	x	x	-	-	-	x	x
DaMaRO	x	-	-	-	x	x	-	-
Open-Exeter	x	x	x	x	x	x	x	x
Total = 6	3	3	4	1	4	4	5	5

Abb. 6: Bibliothekarische Aufgaben in den Projekten des Jisc-RDM-Programms 2011–13¹¹⁸

In einer vergleichenden Analyse von sechs JISC-Projekten wurde zu eruieren versucht, welche bibliothekarischen Aufgaben beim FDM zu erbringen sind (siehe Abb. 6). Zu den wichtigsten Aufgabe gehören ein umfassender Service bzgl. des FDM und entsprechende Trainingsprogramme, wobei die Bibliothek in der Hälfte der Fälle auch die Projektleitung übernommen hat. Weiter war das Management von FDRs und die Entwicklung von Metadaten-Sets und Policies in bibliothekarischer Hand, konnte aber auch von IT-Personal übernommen werden. In 50% der Fälle sorgte die Bibliothek auch für eine Evaluierung des Bedarfs und den sich ergebenden

115 Siehe Joint Information Systems Committee (o. J.): Jisc strategy 2013-16

116 Siehe Digital Curation Centre (o. J.): History of the DCC

117 Siehe Southall; Jisc (2013): Managing Research Data Programme 2011-13

118 Grafik aus: Schmidt (2013): Forschungsdatenmanagement und Bibliotheken, S. 271

den Anforderungen. DMPs wurden eher vom Forschungservice angeboten.¹¹⁹

Im Konkreten hat das Data Management Rollout at Oxford (DaMaRO) Projekt an der University of Oxford versucht die über die Jahre angestrebten FDM-Services in einer ganzheitlichen Infrastruktur zusammenzufassen.¹²⁰ Zentral sind dabei die Systeme DataBank und DataFinder:

„DataBank will become the central institutional archive for research data produced by researchers at Oxford, whereas DataFinder will act as a catalogue or registry of such data – enabling it to be discovered, and, if the appropriate permissions are in place, to be accessed and re-used.“¹²¹

Gerade der DataFinder bietet durch seine Metadatenfunktionen: DataCite-Pflichtfelder, Harvesting interner und externer Ressourcen (auch bibliografischer) den Forschern Unterstützung, die oft vernachlässigte Pflege von Metadaten zu verbessern.¹²²

Es lassen sich einige praktische Ergebnisse des Projekts festhalten:

- Die Bedürfnisse der Forschenden müssen der Maßstab für die FDM-Infrastruktur sein: “[...] if the researchers do not use the infrastructure in practice, compliance is merely theoretical.“¹²³
- Die FDM-Services benötigen eine adäquate Öffentlichkeitsarbeit bei den Forschern: „It seems that researchers tend to pay more attention to developments in their disciplines than to what their institutions are offering them.“¹²⁴

119 Siehe ebenda, S. 271 f.

120 Siehe Wilson (2011): Welcome to the DaMaRO website

121 Wilson; Jeffreys (2013): Towards a Unified University Infrastructure, S. 238

122 Siehe ebenda, S. 238

123 Ebenda, S. 245

124 Ebenda, S. 245

- Infrastrukturen sollten eher modular aufgebaut sein, um adaptiv zu zukünftigen Technologien und Best Practices zu sein.¹²⁵ So könnten sie – in Pionierprojekten entwickelt - eine gute Grundlage für nachfolgende Projekte bilden.

¹²⁵ Vgl. ebenda, S. 245

4 Erhebungen

4.1 Befragung 2012

Im Frühjahr 2012 wurde vom ZAK-CODIGT und dem SCC eine Befragung zur Speicherung und LZA von Forschungsdaten unter Beteiligung von 25 Instituten bzw. OEs (Organisationseinheit) durchgeführt. Die Präsentation der Ergebnisse erfolgt in Auswahl im Hinblick auf die Zielsetzung der vorliegenden Arbeit. Die Grafiken stellen die absoluten Originalwerte dar, wohingegen im Text ggf. gerundete Prozentwerte errechnet wurden oder nur eine Rangfolge der Antworten angegeben wird. Es waren mitunter Mehrfachantworten pro Forscher möglich.

Die Menge der Forschungsdaten ist heterogen: Office-Daten (z.B. Word, PDF, auch LaTeX für Publikationen) gefolgt von wissenschaftlichen Daten in eigenen, teils proprietären Programmen (z.B. SPSS) auch in Verbindung mit technischen Geräten oder als Simulationsdaten (siehe Abb. 7).

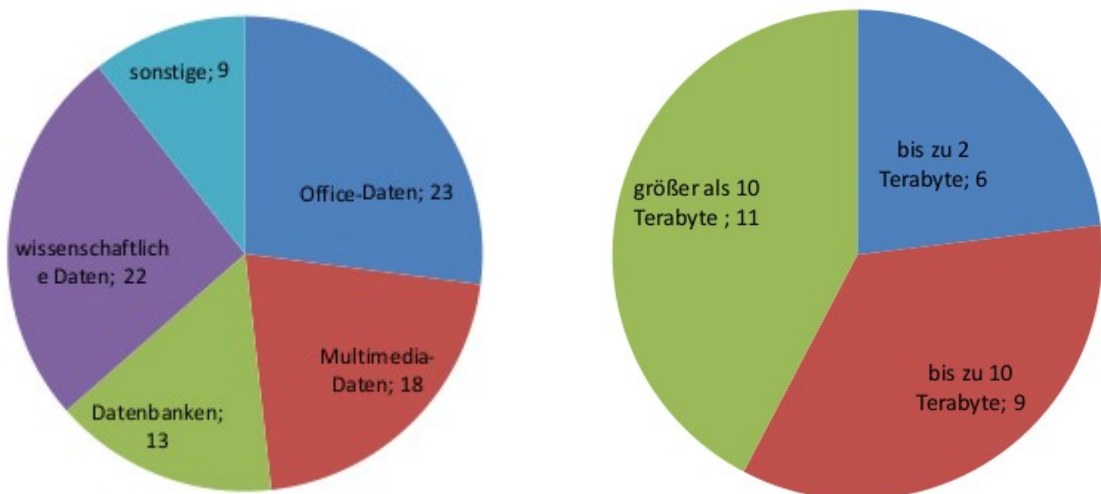


Abb. 7 (links): Datentypen und Datenformate¹²⁶

Abb. 8 (rechts): Gesamtmenge der gespeicherten Daten in einer OE¹²⁷

¹²⁶ Grafik aus: ZAK-CODIGT (2012): dLZA von Forschungsdaten am KIT. Auswertungsdiagramme, S. 2 Nr. 1.1

¹²⁷ Grafik aus: ebenda, S. 11 Nr. 4.1

Weiter gibt es Multimedia-Daten (z.B. Audio- und Video-Dateien, Websites) und Datenbanken. Sonstige Daten können z.B. Programme sein, die selbst Forschungsdaten darstellen.¹²⁸ Die Größe der gespeicherten Daten geht oft in den TB-Bereich (siehe Abb. 8) - mit steigender Tendenz - , wobei nicht alle Daten auch längerfristig speicherungswürdig sind.¹²⁹

37% der Befragten arbeiten mit personenbezogenen Daten und 23% der Forschungsdaten sind patent- oder lizenzrechtlich geschützt, auch durch die Verbindung mit Industriepartnern. Allgemein können Zugangsbeschränkungen, Verschlüsselung und Schutzfristen zum Einsatz kommen.¹³⁰

Für 76% der Wissenschaftler gibt es keine standardisierten Metadaten-Modelle in ihrem Fachbereich (national, international).¹³¹ Bei 40% der Befragten werden Forschungsdaten mit technischen Metadaten versehen, die auch automatisch erstellt werden können.¹³² Bei 48% kommen inhaltliche Metadaten in Form einer Beschreibung durch den Dateinamen vor, 44% arbeiten mit detaillierten Projektberichten und der Rest nutzt andere Beschreibungsmodelle.¹³³

Der Großteil der Forscher (65%) speichert seine Daten am Institut, 30% am SCC (auch LSDF) und der Rest an sonstigen Orten. Die Speicherung kann an mehreren der genannten Orten erfolgen oder durch das Projekt bestimmt sein.¹³⁴ Fast alle Befragten nutzen dafür eine zentrale Datensammelstelle bzw. Datenablage (z.B. Server auf Windows- oder

128 Siehe ZAK-CODIGT (2012): dLZA von Forschungsdaten am KIT. Auswertungssynopse, Nr. 1.1

129 Vgl. auch ebenda, Nr. 4.2

130 Siehe ebenda, Nr. 3.3

131 Siehe ebenda, Nr. 2.7b

132 Siehe ebenda, Nr. 2.7a

133 Siehe ebenda, Nr. 2.8a

134 Siehe ebenda, Nr. 2.10

Unix-Basis), z.T. mit Backup-Systemen für das ganze Institut oder für Projekt- / Forschungsgruppen.¹³⁵

Die Systematisierung erfolgt nach Projekten, Personen oder thematisch, vereinzelt auch chronologisch, als Mischformen oder anders (z.B. projektübergreifend).¹³⁶ Mehr als die Hälfte der Befragten nutzen kein Dokumenten- bzw. Informationsmanagementsystem, bei den übrigen kommen Sharepoint oder Versionsverwaltungen (Subversion, Git), Wikis, CMS-Systeme oder E-Groupware zum Einsatz.¹³⁷

Einheitliche Policies für das Datenmanagement werden bei 29% der OEs angewendet, können aber zusätzlich in Arbeit sein (z.B. ISO 9001: Qualitätsmanagementsysteme – Anforderungen) oder man orientiert sich an der guten wissenschaftlichen Praxis. Es kommen jedoch Empfehlungen wie strukturierte Dateiablage auf OE-internen Servern mit Backup beim SCC zum Tragen. Weiter existieren bisweilen Vorgaben, dass Mitarbeiter bei Verlassen der OE ihre Forschungsdaten übergeben müssen oder es werden disziplinäre FDRs wie GESIS benutzt. Innerhalb einiger Forschungsgruppen werden fachspezifische Datenstandards (z.B. SEED - Standard for Exchange of Earthquake Data) verwendet. Die Finanzierung längerfristiger Datenspeicherung kann ein Problem darstellen.¹³⁸ Die Einführung der Mitarbeiter in die Policy wird größtenteils nach Bedarf / individuell durchgeführt,¹³⁹ die Einhaltung der Policy erfolgt meist auf Basis der Freiwilligkeit.¹⁴⁰

36% geben an, dass eine Nachnutzung der Daten durch Dritte möglich sei. Für weitere 36% trifft das nur teilw. zu; Probleme können spezielles Kontextwissen, uneinheitliche Strukturierung oder proprietäre Dateiforma-

135 Siehe ebenda, Nr. 2.3a

136 Siehe ebenda, Nr. 2.3b

137 Siehe ebenda, Nr. 2.4

138 Siehe ebenda, Nr. 2.1; Vgl. auch ebenda, Nr. 2.5 und 2.6

139 Siehe ebenda, Nr. 2.2a

140 Siehe ebenda, Nr. 2.2b

te sein. 16% verneinen eine mögliche Nachnutzung.¹⁴¹ Die Wichtigkeit einer Nachnutzung der Forschungsdaten für bestimmte Zielgruppen wird mit 100% für die eigene Institution, mit 76% für die Fach-Community und mit 44% für andere Nutzergruppen (interdisziplinär) bemessen, zentrale Daten können über Veröffentlichung geteilt werden.¹⁴²

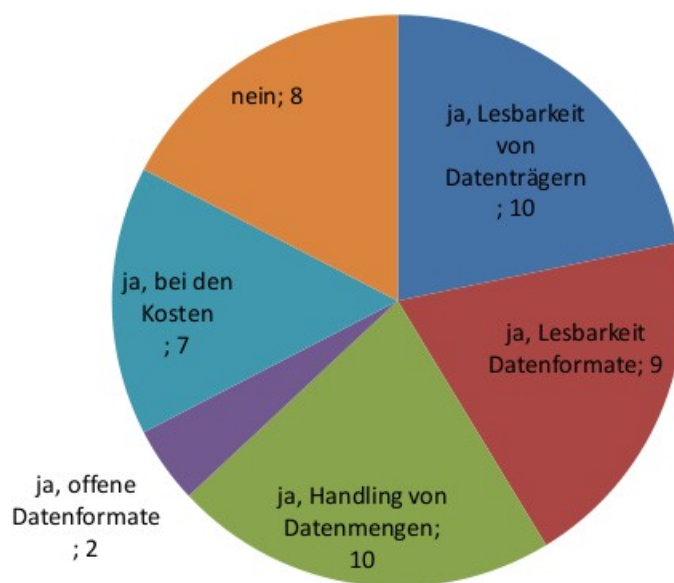


Abb. 9: Erwartete Probleme bei der LZA¹⁴³

Die erwarteten Probleme bei einer LZA verteilen sich auf (siehe Abb. 9): Lesbarkeit der Datenträger (22%), Handling von Datenmengen (22%), Lesbarkeit der Datenformate (20%), Kosten (15%) und offene Dateiformate (4%). Der Rest sieht keine Probleme.¹⁴⁴

¹⁴¹ Siehe ebenda, Nr. 2.9

¹⁴² Siehe ebenda, Nr. 4.3

¹⁴³ Grafik aus: ZAK-CODIGT (2012): dLZA von Forschungsdaten am KIT. Auswertungsdigramme, S. 13 Nr. 4.5

¹⁴⁴ Siehe ZAK-CODIGT (2012): dLZA von Forschungsdaten am KIT. Auswertungssynopse, Nr. 4.5; Vgl. auch ebenda, Nr. 2.11

88% der Forscher sind an Informationen bzw. Best-Practice-Hinweisen zu FDM oder dLZA interessiert und wünschen sich entsprechende Dienstleistungen vom KIT.¹⁴⁵ Vor allem LZA oder deren Optimierung (Ausbau LSDF, Benutzerfreundlichkeit, Flexibilität, Leistungsfähigkeit, Interoperabilität) sind den Forschern ein Anliegen, da fehlende personelle Kapazitäten eine eigene Archivierung oftmals verhindern. Wichtig ist den Forschern auch ein gemeinsamer Datenzugriff (data sharing), evtl. als zentrales Datenbanksystem mit Webinterface oder FTP, auch mit Authentifizierung und Autorisierungsmechanismen. Es werden zudem Dienste zur Veröffentlichung als zitierbare Publikation (z.B. über DOI) gewünscht.¹⁴⁶ Vereinzelt wird auf den Stand der LSDF-Policy oder den Bezug einer finanziellen Ersparnis hingewiesen.¹⁴⁷

4.2 Befragung 2013

4.2.1 Methodik

Zur empirischen Fundierung der Forschungsthematik wurde am KIT vom 20.11.2013 – 16.12.2013 eine Befragung durchgeführt. Diese Untersuchung ist eine Fortführung der im vorangegangenen Kapitel erwähnten Umfrage unter Beteiligung der KIT-Bibliothek und wurde aus Gründen der Handhabbarkeit in zwei Teile (FDR / LZA) getrennt.

65 Institute wurden per E-Mail (inkl. PDF-Formular) angeschrieben, die hauptsächlich aus aktiven Interessenten der vorausgegangenen Befragung bestanden. Zusätzlich sollten noch nicht angesprochene, potentielle Kandidaten berücksichtigt werden und bei entsprechendem Interesse wurden exemplarisch vier persönliche Interviews geführt (siehe Kapitel 4.2.3).

¹⁴⁵ Siehe ebenda, Nr. 5.1 und 5.2a

¹⁴⁶ Siehe ebenda, Nr. 5.2b

¹⁴⁷ Siehe ebenda, Nr. 5.2a

Zehn Forscher haben die Fragebögen beantwortet, damit beträgt die Rücklaufquote 16%.

Die Auswertung der Umfrage und der Interviews erfolgt qualitativ unter Berücksichtigung von Anonymisierungsmaßnahmen. Da das Thema FDRs am KIT noch nicht untersucht wurde, wurde diese Methodik gewählt, um die Situation zu explorieren und Einzelfällen im Sinne von Fallstudien genügend Platz einzuräumen. Die Leitfragen der Interviews orientieren sich an dem Teil 1 der aktuellen Umfrage, den in Kapitel 2 aufgeführten Themenbereichen und der Checkliste aus dem WissGrid-Projekt, wurden aber auf den jeweiligen Kandidaten nochmals zugeschnitten. Die Ergebnisse der Interviews spiegeln die subjektiven Erfahrungen der Forscher (für ihren Fachbereich, für ihr FDR) wider. Auf eine durchgängige Beschreibung im Konjunktiv wurde verzichtet.

Die Ergebnisse der Umfrage und der Interviews werden im Folgenden dargestellt. Eine Analyse und Diskussion erfolgt im Kapitel 5.

4.2.2 Umfrageergebnisse

1. **Disziplin:** Welchen Fachrichtungen ordnen Sie Ihre Forschungstätigkeit und Ihre Forschungsdaten zu?

Die wissenschaftlichen Disziplinen verteilen sich auf Naturwissenschaften, Ingenieurwissenschaften und Schnittstellen zu Geisteswissenschaften.

- Analytische und synthetische Chemie
- Biowissenschaften: Entwicklungsbiologie, Zellbiologie, Molekularbiologie
- Physik, mit Schwerpunkt Astroteilchenphysik
- Geoinformatik, Geodäsie, Geowissenschaften
- Meteorologie, Atmosphärenforschung

- Physik, Ingenieurwesen; Strömungsmechanik; experimentell und numerisch
- Datenanalyse und Bildverarbeitung mit Anwendungen in Biologie, Chemie, Medizintechnik, Energiesystemen und Maschinenbau
- Energieinformatik, Energiedatenmanagement, Datenlebenszyklusanalyse
- Biotechnologie
- Architektur, Archäologie, Cultural Heritage, Denkmalpflege, Denkmalschutz

2. FDR-Existenz: Nutzen / Betreiben Sie selbst Forschungsdatenrepositorien (FDRs) in Ihrem eigenen Projekt / Institut, die online (mit oder ohne Zugangsbeschränkung) zugänglich sind?

Vier der angesprochenen Forscher betreiben (noch) kein eigenes FDR in ihrem Institut. Drei arbeiten mit lokalen Speichersystemen (auch mit der LSDF) oder Ordnerstrukturen an der Schwelle zu einem FDR. Weitere drei nutzen eigene FDRs, die teilw. schon in re3data.org nachgewiesen sind (chemotion, KCDC, Aktuelle Wetterwerte deutscher Stationen; Wolkenatlas; Wetter, Wolken, Klima).

3. Fremd-FDR

Nutzen Sie FDRs anderer Forschungsgruppen oder Forschungseinrichtungen? Wenn ja, ...

3.1 -Nutzen: Nutzen Sie die FDRs produktiv und/oder rezeptiv?

Eine weder produktive noch rezeptive Nutzung von Fremd-FDRs geben drei Forscher an. Ein Forscher nutzt andere FDRs produktiv, einer eher rezeptiv. Drei Forscher geben eine generelle Nutzung an (Bsp. produktiv:

SourceForge, rezeptiv: GenBank¹⁴⁸). Ohne Antwort verbleiben zwei Befragte.

3.2 -Auffinden: Wie gelangen Sie zu FDRs anderer Forschungsgruppen oder Forschungseinrichtungen? Kann Ihnen re3data.org dabei Hilfe bieten?

Kein Befragter gibt re3data.org als Hilfsmittel an, wenn auch ein Forscher der Nachweisdatenbank das Potential zuspricht, sonst nicht auffindbare Daten zugänglich zu machen. Fachspezifische FDRs sind entweder bekannt (1) oder werden über traditionelle Wege wie Publikationen (2) oder Austausch mit Kollegen (auch Konferenzen) ermittelt (2). Es können auch Webportale und Suchmaschinen (auch GoogleScholar) zum Einsatz kommen (2), wobei sich einige eher für Publikationsserver (arXiv.org, PubMed) als für FDRs interessieren (2). Insgesamt geben sieben der zehn Forscher eine Antwort, wobei Mehrfachantworten pro Forscher möglich waren.

3.3 -Nutzen interdisziplinär: Nutzen Sie in transdisziplinärer Weise auch FDRs anderer Fachrichtungen? Haben Sie Beispiele?

Sieben Befragte geben keine Nutzung fachfremder FDRs an, einer davon evtl. mit anderem thematischem Fokus. Ein Forscher aus dem Bereich Meteorologie und Atmosphärenforschung gibt die Nutzung des National Oceanic and Atmospheric Administration (NOAA) Geophysical Data Center an. Zwei Befragte haben keine Antwort gegeben.

3.4 -Nachweis: Gibt es FDRs, die Ihrer Meinung nach in re3data.org erschlossen und verzeichnet sein sollten, es aber bislang noch nicht sind? Haben Sie Beispiele?

148 Ergänzt aus den Antworten zur Frage 2 2013

Von drei Forschern werden keine FDRs genannt, die in re3data.org aufgenommen werden sollen. Weitere drei sprechen sich für die Verzeichnung von einzelnen FDRs aus: Johns Hopkins Turbulence Database (JHTDB),¹⁴⁹ European Research Community on Flow, Turbulence and Combustion (ERCOFTAC),¹⁵⁰ icfd database, AGARD turbulence database; Tropical Rainfall Measuring Mission,¹⁵¹ wobei einer davon auch eine Verzeichnung der LDSF anregt. Vier Forscher haben keine Antwort gegeben.

4 FDR-Erschließung: Haben Sie Interesse, Ihre eigenen FDRs für re3data.org zu erschließen bzw. erschließen zu lassen? Wenn ja ...

4.1 -Kriterien: Was sind für Sie die fünf wichtigsten Kriterien zur Beschreibung eines FDRs?

Von den Befragten beziehen sich fünf auf technische Aspekte des FDRs wie Verfügbarkeit (3), Performance (2), Benutzerfreundlichkeit (2) Suchfunktionen (1) sowie Verlinkung im Internet (1). Vier Forscher sprechen sich für Informationen zu rechtlichen Aspekten aus, davon zwei für Datenschutz und Nutzungsrechte und zwei für entsprechende Authentifizierungs- / Autorisierungskontrollen. Vier Forschern sind Metadaten bzw. Dokumentation ein Anliegen, z.B Erhebungsmethodik, Datenformate, -typ, Grundlegendes (Verfasser, Datum) oder der FDR-Aufbau. Drei Forschern ist die Datenqualität eines FDRs wichtig, z.B. im Sinne einer Qualitätssicherung (peer reviewed) oder jeweils die Aktualität oder Vollständigkeit der Daten. Je ein Forscher spricht sich für die Nennung von Ansprechpartnern und für eine LZA aus. Vier Forscher haben keine Antworten gegeben, Mehrfachantworten pro Forscher waren möglich.

149 <http://turbulence.pha.jhu.edu/>

150 <http://www.ercoftac.org/>

151 <http://trmm.gsfc.nasa.gov/>

4.2 Unterstützung: Kann Sie unser inzwischen routiniertes Erschließungsteam dabei unterstützen?

Vier Forscher schließen eine zukünftige Unterstützung nicht aus (z.B. Verschlagwortung) oder wurden bereits unterstützt (1). Ein Forscher hat keinen Bedarf und vier haben keine Antwort angegeben.

5 Beratung: Wünschen Sie Beratung beim Aufbau eigener FDRs?

Drei Wissenschaftler wünschen eine Beratung: genereller Aufbau (auch Digitalisierung) (1), in Bezug auf die Datenmenge (1) oder ein Review zur Verbesserung des bestehenden FDRs (1); einer hat sie schon in Anspruch genommen. Drei Forscher haben momentan keinen Bedarf und drei haben keine Antworten gegeben.

4.2.3 Interview 1

- Analytische und synthetische Chemie: chemotion¹⁵²

Im Fachbereich kommen Forschungsdaten in Form von Strukturfiles (Molfile) vor, die in einer Nomenklatur erfasst werden. Für die Datensätze der Strukturen werden verschiedene proprietäre Programme benutzt (für jede Messmethode), die eigene Datenformate erzeugen. Als Austauschformat hat sich JCAMP etabliert, welches aber Informationsverluste zum Originalformat beinhaltet. Interessant wäre für die Forscherin eine zusätzliche Emulation der benutzten Software, um der erwähnten Format-, Geräte- und Softwarevielfalt im Hinblick auf eine LZA zu begegnen. Die Daten werden, wenn digital, auf institutseigenen Servern in Form von Datenbanken gespeichert. Die Dateigrößen sind bis auf die Kristallographie in der Disziplin eher als klein einzuschätzen. Metadaten werden teils automatisch erzeugt.

152 <http://www.chemotion.net/>

Gründe für die Entwicklung des FDR chemotion waren Lücken in der chemischen Dokumentation, die eine sinnvolle Nachnutzung erschweren. Zwar gibt es vergleichbare FDRs (z.B. ChemSpider), die aber nicht alle gewünschten Funktionen leisten. Wichtige Aspekte sind ein schnelles und einfaches Einpflegen der Forschungsdaten sowie eine strukturbezogene Suchbarkeit. Nicht zuletzt will man eine Möglichkeit anbieten, Forschungsdaten, die sonst wieder verschwinden würden (z.B. im Zuge von Diplomarbeiten), zu erhalten und zur Verbreitung von Forschungsdaten unabhängig von einer Publikation anregen; oder die Daten vor eine Publikation veröffentlichen (structure first, vgl. Bermuda Principles)

Technisch realisiert wurde das FDR durch ein Web-Application-Framework (Ruby on Rails, JavaScript), das als Open Source auf GitHub veröffentlicht ist und auch von Dritten nachgenutzt werden kann. Die momentane Speicherung in der Amazon Cloud soll zugunsten einer Ablage im LSDF aufgegeben werden, um eine bessere Datenhoheit zu erlangen und Schritte für die LZA einzuleiten.

Wichtig war der befragten Wissenschaftlerin ein Nachweis in re3data.org, um KIT-intern und nach außen Anerkennung zu erlangen und Grundsteine für eine geplante Zertifizierung zu legen. Die Qualitätssicherung der Forschungsdaten wird momentan aufgrund der Überschaubarkeit durch ein internes Review-System gewährleistet. Zukünftig wird dieses z.B. durch externe Gutachter ausgebaut. Es ist eine Integration eines elektronischen Laborjournals an das FDR geplant.

Die KIT-Bibliothek hat bzgl. des FDR mehrere Dienstleistungen erbracht. Durch das re3data.org-Schema ließen sich Anforderungen an ein FDR ableiten, die den Betreibern von chemotion Hinweise gaben um z.B. eine Priorisierung der Eigenschaften eines FDR vorzunehmen: Mindestanforderungen, optionale bzw. längerfristige Features. Weiter hat die KIT-Bibliothek die Rolle eines Vermittlers eingenommen, um Unterstützungen KIT-intern an die relevanten Einrichtungen weiterzuleiten (SCC,

ZAK-CODIGT) oder extern über die den Vorgang der DOI-Adressierung durch die TIB Hannover aufgeklärt.

Wichtig ist der Forscherin für den weiteren Ausbau von chemotion eine Unterstützung bei der Software-Entwicklung.

4.2.4 Interview 2

- Physik, mit Schwerpunkt Astroteilchenphysik: KCDC¹⁵³

Nach Ende des KASCADE Experiments (KARlsruhe Shower Core and Array DEtector) wurde die Idee geboren, die nahezu abgeschlossene und bearbeitete Datenbasis in Form eines FDR bereitzustellen.

Aufgrund der allgemeinen Entwicklung in der Wissenschaft Forschungsdaten zu veröffentlichen, möchten die beteiligten Forscher im Bereich der Astroteilchenphysik Pionierarbeit leisten. Unter Berufung auf die Helmholtz-Policy und die Berliner Erklärung sollen die von der öffentlichen Hand finanzierten Daten als Open Data der Allgemeinheit zur Verfügung stehen.

Technisch wird ein Open Source Web-Framework benutzt, das in einem weiteren Schritt auch zur Nachnutzung für vergleichbare Projekte (KIT, Dritte) freigegeben wird. Die mehrere GB umfassenden Datenpakete können durch Parameter (Cuts) selektiert und als Textdateien heruntergeladen werden. Man hat sich gegenwärtig gegenüber etablierten Formaten (Root-Files) bewusst für ein menschenlesbares Dateiformat entschieden, um eine programmunabhängige Lesbarkeit zu garantieren. Je nach Feedback der Zielgruppen ist die Implementierung weiterer Formate (Root-Files, HDF), die auch Kompressionen unterstützen, geplant. Neben der für die Entwicklung notwendigen Vorhaltung der Daten am Institut ist langfristig eine LZA in Kooperation mit dem SCC angedacht.

153 <https://kcdc.ikp.kit.edu/>

Die Qualitätssicherung erfolgt projektintern anhand disziplinspezifischer Kriterien, wobei auch eine externe Zertifizierung diskutiert wird. Rechtlich hat man sich für eine EULA (End-User License Agreement), wie sie bei Webportalen in der Industrie verwendet wird, entschieden, die in Kooperation mit der Rechtsabteilung des KIT entwickelt wurde. So möchte man sich eine Modifizierbarkeit (z.B. bzgl. der aktuell zu zitierenden Publikation) vorbehalten und im Sinne guter wissenschaftlicher Praxis gegen externe Ansprüche oder Rufschädigungen absichern.

Eine Zitierbarkeit der Daten ist nach Angaben des Befragten durch die Cuts gewährleistet, die bei gleichen Parametern die gleichen Ergebnisse liefern. Eine erweiterte persistente Adressierung ist neben Bereitstellung einer API (Application Programming Interface) in Verbindung mit XML-Files im Gespräch. Neben internen, technischen Metadaten werden manuelle Dokumentationen in Form von weboptimierten Infoboxen und Größenbeschreibungen verwendet. Die Betreiber des FDR haben es sich zusätzlich zur Aufgabe gemacht, die Daten für Schüler aufzubereiten und Lehrmaterialien bereitzustellen.

Technische Dienstleistungen außerhalb des Instituts wurden für die Entwicklung des FDR nicht in Anspruch genommen, da durch die Datenintensität der Physik bereits eine Expertise vorhanden ist. Es besteht aber eine Kooperation mit dem SCC, die man unter Umständen in Bezug auf Webtechnologien vertiefen kann. Die KIT-Bibliothek konnte nach der Entwicklung des FDR den Nachweis in re3data.org und die Vernetzung in der FDR-Community im Sinne eines Best Practice beisteuern. Im Bereich der Metadaten und der Optimierung bzgl. des Information-Retrievals wird das größte Dienstleistungspotential gesehen, um das FDR auch weiterhin aktuell und benutzerfreundlich zu gestalten.

4.2.5 Interview 3

- Energieinformatik, Energiedatenmanagement, Datenlebenszyklusanalyse

Der Fachbereich arbeitet vorwiegend mit Datenbanken (MySQL, NoSQL), wobei auch tabellarisch strukturierte Daten (Excel, CSV) oder Textdateien vorkommen. Da Energieverbrauchsdaten (z.B. in einem Smarthome) Nutzungsmuster erkennen lassen und so einen Personenbezug haben, werden datenschutzrechtliche Bereiche berührt und Maßnahmen der Anonymisierung oder die Vernichtung der Daten durchgeführt. Weiter wird auch mit lizenzrechtlich geschützten Daten wie georeferenzierten Luftbildern zur Dachflächenabschätzung für Photovoltaik-Anlagen gearbeitet. Die Forschungsdaten sind bis auf Simulationsdaten wenig standardisiert; ihnen werden teils manuell Metadaten (gemessene Geräte, Zeitpunkte) hinzugefügt. Die wissenschaftliche Veröffentlichung erfolgt klassisch als Auszug im Anhang von Publikationen, wobei der Wunsch nach besserer Referenzierbarkeit der Forschungsdaten vorhanden ist. Die Daten werden je nach Projektpolicy am Institut / der Firma vorgehalten und für zehn Jahre gesichert. Eine Kooperation mit dem SCC besteht und das Interesse für eine LZA ist vorhanden. Durch den Personenbezug existieren FDRs oft nur für den projektinternen Gebrauch oder in Kollaborationen (vgl. shared domain). Zusätzlich verhindern finanzierende Industriepartner eine systematische Veröffentlichung der Daten und auch in der Fach-Community ist das data sharing, bis auf ältere bzw. abgeschlossene Projekte oder einzelne Anfragen von Forschern, wenig verbreitet.

Technische Dienstleistungen werden durch die vorhandene Expertise im Datenhandling selbst erbracht oder durch das SCC, das auch beratend tätig ist, vorgenommen. Mit re3data.org hat der Befragte noch nicht konkret gearbeitet, könnte aber das Potential haben datenschutzrechtlich unbe-

denkliche Forschungsdaten (z.B. Photovoltaikdaten, die sonst simuliert werden müssten oder Nutzungsdaten von Elektrofahrzeugen, Windkraftanlagen) zugänglich zu machen.

4.2.6 Interview 4

- Architektur, Archäologie, Cultural Heritage, Denkmalpflege, Denkmalschutz¹⁵⁴

Die Photogrammetrie verbindet fotografische Aufnahmen von Objekten mit einer metrischen Erfassung durch so genannte Punktmarken / Passpunkte. Eine räumliche Dokumentation wird u.a. durch stereoskopische Erfassung mittels Bildpaaren möglich, die bspw. in digitalen 3D-Modellen gespeichert und visualisiert werden kann. Die Fach-Community ist eher klein, wobei durch den Themenbereich des Cultural Heritage stark interdisziplinär gearbeitet wird und eine große Bedeutung der Arbeit für die Öffentlichkeit besteht (Aufnahmen von Weltkulturerbe, Google Earth inkl. 3D-Gebäudemodelle etc.). Gleichzeitig haben die geleisteten Arbeiten einen Unikat-Charakter, da sie z.B. historisch bedeutsame Objekte dokumentieren, die heute schon einen wachsenden Verfall aufweisen. Ein weiterer Motor für die erkannte Notwendigkeit eines FDM und einer Archivierung ist der Umstand, dass spezialisiertes Know-How, welches zur Interpretation der Daten erforderlich ist, durch altersbedingtes Ausscheiden der Mitarbeiter verloren zu gehen droht.

Im konkreten Fall geht es um Forschungsdaten aus Projekten, welche zwischen 1300 v. Chr. - 1910 n. Chr. entstandene Baudenkmäler und archäologische Objekte (z.B. Pompeji, Augustus-Forum (Rom), Tyrens, Schloss Heidelberg etc.) dokumentieren. Die Arbeiten wurden von 1968 bis 2008 ausgeführt und sollen nun neben fachverwandten Zielgruppen (auch Allge-

154 auch: Photogrammetrie, Fernerkundung, Geoinformatik

meinhistoriker, Journalisten und Geoinformationsdiensten) auch einer breiteren Öffentlichkeit zur Verfügung gestellt werden.

Die Daten liegen in Form mehrerer Tausend fotografischer Glasplatten und Lageskizzen sowie Koordinatenlisten als Medienverbund vor, die digitalisiert, erschlossen und online in einem FDR präsentiert werden sollen. Die Digitalisierung muss aufgrund der Menge und den spezifischen Genauigkeitsanforderungen durch einen externen Dienstleister erfolgen. Die Originale sollen danach im KIT-Archiv verbleiben und nach den üblichen Standards (General International Standard Archival Description, ISAD G sowie mit Rücksicht auf den Strukturstandard Encoded Archival Description, EAD) erschlossen werden. Die Größe der Digitalisate wird auf ein bis zwei TB (z.B. TIFF, JPEG) geschätzt und für die Metadaten werden Excel oder XML-Formate in Betracht gezogen. Bei der Erschließung soll Pionierarbeit geleistet werden, von deren Erkenntnissen auch andere Photogrammetrie-Experten Rückschlüsse auf einen optimalen Workflow und zu verwendende Standards (Kompatibilität mit Dublin Core) ziehen können. Da die Daten zum Zeitpunkt der Erhebung nur für eine projektbezogene Nutzung vorgesehen waren, hat man keine allgemeingültigen Standards verwendet, sondern die Daten nach internen Kriterien systematisiert. Die Daten sind rechtlich zum größten Teil als gemeinfrei einzustufen, da keine exklusiven Nutzungsrechte mit den Kostenträgern vereinbart wurden und die Verwertungsrechte somit bei den Urhebern bzw. dem KIT liegen.

Neben der Präsentation der Forschungsdaten im FDR ist in einem weiteren Schritt eine Visualisierung der Daten in Form fotorealistisch texturierter 3D-Modelle für eine virtuelle Begehung der Objekte angedacht. Zudem soll das Einpflegen aktueller photogrammetrischer Daten (digital born) bzw. eine generelle Harmonisierung mit Neuaufnahmen möglich sein. Da es kein vergleichbares FDR für den Fachbereich gibt, ist den Betreibern eine Nachnutzung auch für Dritte (rezeptive/produktive Nutzung des FDR, Weitergabe der FDR-Software) ein Anliegen. Die Forschungsdaten, die in

wissenschaftlichen Arbeiten publiziert wurden, spiegeln nur Auszüge der ganzen Daten wider. Auch hier soll das FDR die Lücke schließen und Vollständigkeit anstreben. Bzgl. einer Referenzierung der Daten sind persistente Adressierungen (z.B. über URN) im Gespräch. Neben der analogen Archivierung der Originaldaten („dark archive“) wird eine dLZA (SCC, ZAK-CODIGT) der Forschungsdaten angestrebt.

Die Forscher der Projekte haben verschiedene Dienstleistungen des KIT in Anspruch genommen. Es wird intensiv mit dem KIT-Archiv zusammengearbeitet, welches gerade für Fragen der Archivierung der Originale und Digitalisierung eine Expertise besitzt, die beratend genutzt wurde. Die KIT-Bibliothek prüft zusammen mit dem SCC und ZAK-CODIGT Lösungsmöglichkeiten. Durch ihre Erfahrungen mit FDRs können wichtige strategische Entscheidungen (z.B. auch Finanzierungs- und Kostenfragen) unterstützt werden. Eine Erschließung in re3data.org und die damit verbundene Sichtbarmachung des FDR wird auch gewünscht. Der größte Bedarf besteht momentan in einer technischen Unterstützung bzgl. der FDR-Entwicklung sowie auch bzgl. Digitalisierungsdiensten.

5 Analyse

Anhand der Darstellungen aus dem theoretischen und praktischen Teil sowie den Ergebnissen aus den Erhebungen werden im Folgenden eine Analyse vorgenommen und Empfehlungen ausgesprochen. FDR-Aspekte lassen sich nicht einheitlich vom (vorausgehenden) FDM trennen, deshalb wird beides integriert dargestellt. Es lassen sich zuerst allgemeine, fachunabhängige Erkenntnisse festhalten.

Die **Daten** sind laut Frage 2.1 (2012) und den Interviews heterogener Natur, was sich mit theoretischen Erkenntnissen deckt. Gleichzeitig sind die Datenmengen (Frage 4.1, 2012) eher groß. Dies wird aber durch die exemplarischen Fälle der Interviews etwas relativiert, sobald es um ein konkretes FDR geht. Im Bereich der **Metadaten** für Forschungsdaten ist kein umfassender Schluss möglich (Frage 2.7 - 2.8, 2012); nach Meinung des Verfassers sind die Disziplinen dahingehend aber eher autonom. Gerade für eine standardisierte Grunderschließung könnte aber ein Bedarf bestehen (Frage 4.1, 2013). Das Interesse an einer LZA (Frage 5.1 und 5.2a 2012, Interview 1-4) lässt sich auch auf Metadaten der LZA ausweiten.

Forschungsdaten werden bereits in der **KIT-Policy** angesprochen und damit nationale Policies institutionell umgesetzt. Es könnte aber zu einer Ergänzung wie bei der Universität Bielefeld kommen, die in ihrer Resolution, zur Nutzung der universitären FDM-Dienstleistungen aufruft und auf die Veröffentlichung in FDRs und den Nachweis in re3data.org hinweist.¹⁵⁵ Mit den Erfahrungen aus der Umfrage von 2012 (Frage 2.1 - 2.3) könnten auch Maßnahmen zum Datenmanagement konkreter adressiert werden oder Unterstützung bei DMPs angeboten werden (wenn auch bei den Interviews nicht gefordert).

Rechtliche Aspekte sind sowohl bei den Daten selbst (Frage 3.3, 2012) als auch bei FDRs (Interviews 1-3) ein Thema. Zum einen ist die Wahrung

¹⁵⁵ Siehe Universität Bielefeld (2013): Resolution zu Forschungsdatenmanagement

des Datenschutzes für Forscher wichtig, die hauptsächlich mit personenbezogenen Daten arbeiten. Zum anderen spielen lizenzrechtliche Aspekte vornehmlich bei Vorhaben eine Rolle, die durch Industriepartner (mit-)finanziert werden (Interview 3). Für ein FDR werden zukünftig Autorisierungs- und Authentifizierungsmechanismen wichtig sein (Frage 4.1, 2013).

Die von einigen Wissenschaftlern geforderte (Frage 5.2b, 2012) Dienstleistung zur **Referenzierung** (z.B. DOI-Vermittlung) wird bereits von der KIT-Bibliothek angeboten und bereits genutzt (Interview 1). Die **LZA** ist wohl eines der größten zukünftigen Aufgabenfelder im Bereich der Forschungsdaten. Der Bedarf bei KIT-Forschungsgruppen ist deutlich erkennbar (Frage 5.2b, 2012; Interviews 1-4). Es besteht schon eine Kooperation zwischen der KIT-Bibliothek und dem SCC, die zukünftig ausgebaut werden soll. Weiter muss geklärt werden, wie eine optimale Verzahnung von OE-eigenen FDRs im Hinblick auf ein späteres Hosting am SCC (LSDF) auszusehen hat, und die Verantwortlichkeiten müssen auf Dauer festgelegt werden.¹⁵⁶ Bei den in FDRs abgelegten Daten werden adäquate LZA-Methoden erforderlich. Diese können in Zusammenarbeit mit ZAK-CODIGT und re3data.org konzipiert werden.¹⁵⁷

Die Umfrage 2012 (Frage 2.5) lässt noch keinen eindeutigen Trend für den Einsatz von **FDRs** am KIT erkennen. Zwar wird auch in der Frage FDR-Software (Fedora, DSpace) genannt, die **Terminologie** war aber damals noch nicht eindeutig und die Antworten suggerieren eher interne Dokumentenverwaltung. Das zeigt z.B. auch, in welchem Wandel sich das Thema befindet und wie nah technische Systeme mit ihren Funktionalitäten beieinander liegen können. Die neuere Umfrage (2013) zeigt, dass in der Zwischenzeit die Bestrebungen, eigene FDRs auszubauen und zu betreiben, konkreter geworden sind. So sind einige FDRs des KIT (bspw. chemotion, KCDC) bereits online und in re3data.org nachgewiesen.

¹⁵⁶ Gespräch mit Vertreter der Abteilung Scientific Data Management (SCC)

¹⁵⁷ Gespräch mit ZAK-CODIGT Geschäftsführer

Die geringe Rücklaufquote der Umfrage von 2013 kann man auch als Indiz für eine bestehende Unsicherheit gegenüber der Thematik deuten. Zudem gab es offenbar technische Probleme bei der PDF-Funktionalität und einige Adressaten deuteten an, nach der vorherigen Umfrage hätte sich nichts geändert. Weiter werden FDRs am KIT in Frage 3.2 (2013) eher auf klassischen Wegen als mit re3data.org gefunden. Der allgemeine Eindruck aus den Interviews legt nahe, dass viele Forschern nicht genau wissen, welche Dienstleistungen das KIT und speziell die KIT-Bibliothek bzgl. FDM oder FDRs überhaupt anbieten. Durch die Erhebungen wurde die Thematik „Forschungsdaten am KIT“ und entsprechende Anforderungen der Forscher exemplarisch evaluiert. Gleichzeitig kann man solche Umfragen als Maßnahme einer **KIT-internen Öffentlichkeitsarbeit** begreifen, welche z.B. zwei Interviewpartner zur unmittelbaren Inanspruchnahme von FDM-Dienstleistungen motiviert. Solche Maßnahmen sollten in regelmäßigen Abständen durchgeführt und bestenfalls ausgeweitet werden. Dafür sprechen auch positive Erfahrungen an der Universität Oxford (Kapitel 3.3). Einige Forscher sprechen sich in Frage 5 (2013) für eine Beratung aus und ein Forscher hält „gebündelte, fundierte Information“ für nützlich. Die Antworten legen allgemein eine Durchführung von Vorträgen und Schulungen nahe, wie sie auch im Pyramidenmodell angesprochen wird, z.B. unter Einbeziehung der Lehrmaterialien des DCC.

Digitalisierungsdienste, wie sie bei dem Interview 4 (Photogrammetrie) angesprochen wurden, fallen nicht direkt in das FDM und müssen gesondert behandelt werden.

Vergleiche mit anderen nationalen und internationalen Institutionen sind für grundlegende FDM-Maßnahmen oder Einzelprojekte erstrebenswert, eine generelle Übertragbarkeit ist aber nicht möglich. Die Universität Oxford (siehe Kapitel 3.3) und die TU Berlin¹⁵⁸ arbeiten bspw. mit institutionellen FDRs, welche auch Vernetzungen mit externen Ressourcen anstre-

158 Siehe Kuberek (2013): Organisatorisch-technisches Konzept für eine Forschungsdaten-Infrastruktur in der TU Berlin, S. 4

ben. Diese Herangehensweise ist momentan im KIT nicht aktuell, auch wegen der gleichzeitigen Etablierung eines Forschungsinformationssystems. Die DINI fordert einen eher disziplinären Umgang mit Forschungsdaten.¹⁵⁹ Die Praxiserfahrungen am KIT und mit re3data.org haben gezeigt, dass das auch auf FDRs zutrifft. Genauer betrachtet haben beide Strategien viele Gemeinsamkeiten. Die Erfahrungen aus Großbritannien legen nahe, dass wenn disziplinäre Infrastrukturen existieren, diese vorzuziehen sind:

„Forschende verhalten sich gegenüber ihrer Fachgemeinschaft im Normalfall loyaler als gegenüber ihrer Institution, insbesondere, weil sie die Institution eher wechseln als das Fachgebiet, aber auch, weil zentrale Services meist zwar Querschnittsorderungen zu erfüllen bemüht sind, dies aber auch dazu führt, dass spezielle Anforderungen unbeachtet bleiben müssen.“¹⁶⁰

Wenn es diese disziplinären Infrastrukturen nicht gibt, füllt das KIT die Lücke zur Zeit nicht mit einem ganzheitlichen **institutionellen FDR**, sondern bietet Unterstützung bei der Etablierung neuer **disziplinärer FDRs** wie chemotion oder institutionell bzw. projektbezogener FDRs wie KCDC. Interessant ist dahingehend auch ein Blick auf Text-Repositoryn (Dokumentenserver), die ein ähnliches Spannungsverhältnis zwischen Institution und Disziplin besitzen, und bei denen Schnittstellen sowie Vernetzung im Sinne einer Infrastruktur ebenso eine kritische Rolle spielen.¹⁶¹

Festzuhalten ist auch die Tendenz, **Open Source-Systeme** zu verwenden, sowohl in der Theorie (Kapitel 2.3.1) als auch bei den lokalen FDRs (Interview 1 und 2) und bei dem Beispiel aus Oxford (Kapitel 3.3). Eine generelle Nachnutzung der Systeme kann so unter Abwägung eines Kos-

159 Siehe Deutsche Initiative für Netzwerkinformation (2009): Positionspapier Forschungsdaten, S. 7

160 Schmidt (2013): Forschungsdatenmanagement und Bibliotheken, S. 267

161 Vgl. Scholze; Müller (2012): Aufbau und Vernetzung eines Repositoriums, S. 14 f.

ten-Nutzen-Verhältnisses, auch in Form von Teilsystemen, Beachtung finden.

Mit **re3data.org** existiert nicht nur eine Infrastruktur, die den Nachweis und die Vernetzung von FDRs gewährleistet. Aus den sich etablierenden Kriterien (Schema) entwickelt sich auch eine effektive Knowledge Base für FDR-Vorhaben (Interview 1) oder deren Optimierung (Interview 2). Durch „re3data4KIT“¹⁶² können so Best-Practice-Ratschläge für Forscher gegeben werden, die sich mit dem Thema FDR auseinandersetzen. Die Kriterien für die Beschreibung eines FDR (Frage 4.1, 2013) - interpretiert aus der Nutzersicht - korrelieren stark mit dem Icon-System von re3data.org mit Betonung auf technische Aspekte. Auch andere Forschungseinrichtungen können aus dem re3data.org-Schema und den damit verbundenen Qualitätsmerkmalen Hinweise erhalten, ihr FDM in Bezug auf die Etablierung von FDRs zu optimieren.

Daneben nimmt die KIT-Bibliothek lokal die Rolle eines Vermittlers ein, um FDR-Vorhaben zu koordinieren (Interview 1 und 4). Dies hat sich wahrscheinlich nicht ohne Grund auch in der britischen Praxis etabliert,¹⁶³ da Bibliotheken eine informationwissenschaftliche Expertise besitzen.

Wie bei den Forschungsdaten selbst, stellen auch beim FDM und bei FDRs disziplinspezifische Ausrichtungen oftmals die attraktivere Lösung dar. Somit können Folgerungen für ganze **Fachbereiche** nur sehr allgemein ausfallen. Man muss zudem die unterschiedlichen Entwicklungsstufen der Infrastruktur in den einzelnen Fächern berücksichtigen (Bio-, Geo- und Sozialwissenschaften als Vorreiter).¹⁶⁴

Bspw. suggeriert das entsprechende Interview in der (Astro-)physik, dass für datenintensive naturwissenschaftliche Fächer eher weniger technische Dienstleistungen zum Aufbau eines FDRs benötigt werden. Erst für die

162 Vorläufiger Arbeitstitel

163 Siehe Schmidt (2013): Forschungsdatenmanagement und Bibliotheken, S. 260 f.

164 Siehe Deutsche Initiative für Netzwerkinformation (2009): Positionspapier Forschungsdaten, S. 6

LZA sollten konkrete Dienstleistungen in Absprache mit dem SCC, ZAK-CODIGT angeboten werden, da andere informationstechnische Bereiche als bei der Datenhaltung berührt werden. Für Fächer oder Institute, die wenig Erfahrungen mit Datenhandling haben, werden technische Dienstleistungen zum Aufbau eines FDR benötigt. Diese muss von allen Service-OEs in Form von aufeinander abgestimmten Modulen (z.B. FDM-Plan, Bit Streaming, FDR-Baukasten, APIs, dLZA, klassische Archivierung) erbracht werden.

Auf Seiten der Bibliothek könnte der verstärkte Einsatz von Fachreferenten erwogen werden, um der erwähnten, eher disziplinären Ausrichtung der FDM Rechnung zu tragen.¹⁶⁵

Mit wachsender Anzahl an FDR-Vorhaben bilden sich sukzessive weitere fachabhängige Gemeinsamkeiten aus.¹⁶⁶ Zusätzlich bieten sich zwei Varianten an, um die Dienstleistungen dementsprechend zu konzentrieren: (1.) eine Einteilung der interessierten Forschergruppen in verschiedene **Typen**, (2.) eine **Priorisierung** in grundlegende und weitergehende Bedürfnisse.

- existiert FDR? - ja / nein (disziplinär, institutionell, projektbezogen)
- möchte in re3data.org? - ja / nein
- braucht technische / organisatorische Unterstützung?
- benötigt spezielle Beratung (DOI, LZA, DMP, Metadaten ...)?
- benötigt allgemeine Beratung?
- ...

¹⁶⁵ Vgl. Schmidt (2013): Forschungsdatenmanagement und Bibliotheken, S. 270

¹⁶⁶ Vgl. hierzu Neuroth (2012): Langzeitarchivierung von Forschungsdaten

6 Ausblick

Gerade die LZA von FDRs oder allgemein von Forschungsdaten ist ein hochaktuelles Aufgabengebiet, an dem auch am SCC gearbeitet wird. Im Rahmen dieser Arbeit konnten informationstechnologischen Details nur angerissen werden. Man beachte hierzu auch die Projekte LOCKSS (Lots of Copies Keep Stuff Safe) und kopal (Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen).

Ein größer angelegter Vergleich, z.B. mit ähnlichen Einrichtungen wie dem Forschungszentrum Jülich oder mit Forschungsgesellschaften wie der Fraunhofer-Gesellschaft und der Max-Planck-Gesellschaft, auch im internationalen Bereich stellt ein Forschungsdesiderat dar. Zudem können kommerzielle Anbieter wie Google mit seinen Research-Programmen und Diensten wie dem Public Data Explorer sowie WolframAlpha (computational knowledge engine) und Elsevier mit Scopus oder seinem research data service verstärkt als potentielle Servicepartner auftreten. Auch die Rolle der Verbundzentralen wie dem BSZ (Bibliotheksservice-Zentrum Baden-Württemberg) wird in dieser Hinsicht neu überdacht werden müssen.

Gleichzeitig bieten Technologien wie Linked Open Data (LoD) und virtuelle Forschungsumgebungen (VFU),¹⁶⁷ Forschungsinformationssysteme (FIS) in Kombination mit Forschungsdaten in der Zukunft ganz neue Möglichkeiten. Abzuwarten bleibt wie sich diese oder andere Teile der Informationsstruktur in die Thematik einfügen und wie Schnittstellen zu FDRs umgesetzt werden. Auch die Ergebnisse des vor kurzem gestarteten Projekts bwFDM (Forschungsdatenmanagement Baden-Württemberg) können neue Erkenntnisse zu Tage bringen.

¹⁶⁷ Vgl. auch die BW-eLabs

Quellenverzeichnis

- Altenhöner, Reinhard, Hrsg. (2012): Langzeitarchivierung von Forschungsdaten. Standards und disziplinspezifische Lösungen. 1. Aufl. Berlin, Scivero-Verl.
- American Geophysical Union (2013): AGU Publications Data Policy.
URL: <http://publications.agu.org/author-resource-center/publication-policies/data-policy/> (Zugriff: 7.2.2014)
- Aschenbrenner, Andreas; Neuroth, Heike (2011): Forschungsdaten-Repositoryn, in: Büttner, Stephan (Hrsg.): Handbuch Forschungsdatenmanagement. Bad Honnef, Bock & Herchen, S. 101-114
- Australian National Data Service (ANDS) (o. J.): Curation Continuum.
URL: <http://ands.org.au/guides/curation.continuum.html>
(Zugriff: 16.12.2013)
- Blanke, Tobias; Horstmann, Axel; Rauber, Andreas (2013): Bericht über das Symposium zum zehnjährigen Bestehen der Abteilung Forschung & Entwicklung, IV., in: Neuroth u. a. (Hrsg.): Evolution der Informationsinfrastruktur. Forschung und Entwicklung als Kooperation von Bibliothek und Fachwissenschaft. Glückstadt, Werner Hülsbusch, S. 39-68
- Brübach, Nils (2010): Das Referenzmodell OAIS, in: Neuroth, Heike; Oswald, Achim (Hrsg.): Nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Göttingen, Nieders. Staats- u. Univ.-Bibl., Kap. 4:3-4:14
- Büttner, Stephan, Hrsg. (2011): Handbuch Forschungsdatenmanagement. Bad Honnef, Bock & Herchen. URL:
<http://opus.kobv.de/fhpotsdam/volltexte/2011/241/pdf/Handbuch-Forschungsdatenmanagement.pdf> (Zugriff: 04.09.2013)
- Büttner, Stephan; Hobohm, Hans-Christoph; Müller, Lars (2011): Research Data Management, in: Büttner, Stephan (Hrsg.): Handbuch Forschungsdatenmanagement. Bad Honnef, Bock & Herchen, S. 13-24
- Cell (o. J.): Information for Authors. Editorial Policies.
URL: http://www.cell.com/authors#ed_policies (Zugriff: 7.2.2014)
- Corall, Sheila (2012): Roles and responsibilities. Libraries, librarians and data, in: Pryor, Graham (Hrsg.): Managing Research Data. London, Facet Publishing, S. 105-133

- Dallmeier-Tiessen, Sünje (2011): Strategien bei der Veröffentlichung von Forschungsdaten, in: Büttner, Stephan (Hrsg.): Handbuch Forschungsdatenmanagement. Bad Honnef, Bock & Herchen, S. 157-168
- DataCite (o. J.): German National Library of Science and Technology. URL: <http://www.datacite.org/TIB> (Zugriff: 27.1.2014)
- Deutsche Forschungsgemeinschaft (DFG) (2009): Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten. URL: http://www.dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901.pdf (Zugriff: 14.1.2014)
- Deutsche Forschungsgemeinschaft (DFG) (2013): Vorschläge zur Sicherung guter wissenschaftlicher Praxis. Denkschrift. Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“. Weinheim, Wiley-VCH. URL: http://www.dfg.de/download/pdf/dfg_im_profil/re-den_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf (Zugriff 13.1.2014)
- Deutsche Forschungsgemeinschaft (DFG) (o. J.): Sicherung von Forschungsdaten. URL: http://www.dfg.de/foerderung/antragstellung_begutachtung_entscheidung/antragstellende/antragstellung/sicherung_forschungsdaten/ (Zugriff: 6.2.2014)
- Deutsche Initiative für Netzwerkinformation (DINI), Arbeitsgruppe Elektronisches Publizieren (2009): Positionspapier Forschungsdaten, in: DINI-Schriften 10. URN: urn:nbn:de:kobv:11-10098082
- Digital Curation Centre (DCC) (o. J.): DCC Curation Lifecycle Model. URL: <http://www.dcc.ac.uk/resources/curation-lifecycle-model> (Zugriff: 4.2.2014)
- Digital Curation Centre (DCC) (o. J.): General Research Data. URL: <http://www.dcc.ac.uk/resources/subject-areas/general-research-data> (Zugriff: 3.2.2014)
- Digital Curation Centre (DCC) (o. J.): History of the DCC. URL: <http://www.dcc.ac.uk/about-us/history-dcc/history-dcc> (Zugriff: 7.1.2014)
- Dobratz, Susanne; Schoger, Astrid (2010): Grundkonzepte der Vertrauenswürdigkeit und Sicherheit, in: Neuroth, Heike; Oßwald, Achim (Hrsg.): Nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Göttingen, Nieders. Staats- u. Univ.-Bibl., Kap. 5:2-5:8

- Dryad (2013): Joint Data Archiving Policy (JDAP). URL: <http://datadryad.org/pages/jdap> (Zugriff: 7.2.2014)
- European Heads of Research Councils (EUROHORC); European Science Foundation (ESF) (2008): The EUROHORCs and ESF Vision on a Globally Competitive ERA and their Road Map for Actions to Help Build It, in: Science Policy Briefing 33. URL: http://www.esf.org/fileadmin/Public_documents/Publications/SPB33_ERARoadMap.pdf (Zugriff 1.12.2013)
- Funk, Stefan E. (2010): Digitale Objekte und Formate, in: Neuroth, Heike; Oßwald, Achim (Hrsg.): Nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Göttingen, Nieders. Staats- u. Univ.-Bibl., Kap. 7:3-7:8
- Goebelbecker, Hans-Jürgen (2010): „Nie war er so wertvoll wie heute.“ Der Primärbericht.
- Hein, Stefan (2012): Metadaten für die Langzeitarchivierung, in: Altenhöner, Reinhard (Hrsg.): Langzeitarchivierung von Forschungsdaten. Standards und disziplinspezifische Lösungen. Berlin, Scivero-Verl., S. 87-109
- Helmholtz-Gemeinschaft (2013): Open Access. Offener Zugang zu Forschungsdaten. URL: <http://oa.helmholtz.de/index.php?id=63> (Zugriff: 6.2.2014)
- Higgins, Sarah (2012): The Lifecycle of Data Management, in: Pryor, Graham (Hrsg.): Managing Research Data. London, Facet Publishing, S. 17-45
- Von der Hude, Nicole (2012): Persistent Identifier. Versionierung, Adressierung und Referenzierung, in: Altenhöner, Reinhard (Hrsg.): Langzeitarchivierung von Forschungsdaten. Standards und disziplinspezifische Lösungen. Berlin, Scivero-Verl., S. 129-135
- Huschka, Denis (Hrsg.) (2011): Datenmanagement und Data Sharing. Erfahrungen in den Sozial- und Wirtschaftswissenschaften, in: Büttner, Stephan (Hrsg.): Handbuch Forschungsdatenmanagement. Bad Honnef, Bock & Herchen, S. 35-48
- International DOI Foundation (o. J.): DOI Handbook. Introduction. URL: http://www.doi.org/doi_handbook/1_Introduction.html (Zugriff: 27.1.2014)

- Jensen, Uwe; Katsanidou, Alexia; Zenk-Möltgen, Wolfgang (2011): Meta-daten und Standards, in: Büttner, Stephan (Hrsg.): Handbuch Forschungsdatenmanagement. Bad Honnef, Bock & Herchen, S. 83-100
- Joint Information Systems Committee (Jisc) (o. J.): Jisc strategy 2013-16. URL: <http://www.jisc.ac.uk/reports/jisc-strategy-2013-16> (Zugriff: 7.1.2014)
- Jones, Catherine; Darby, Robert; Gilbert, Linda; Lambert, Simon (2008): Report of the Subject and Institutional Repositories Interactions study., in: CMJ 7, H. 9
- Karlsruher Institut für Technologie (KIT) (2010): Regeln zur Sicherung guter wissenschaftlicher Praxis im Karlsruher Institut für Technologie (KIT). URL: http://www.kit.edu/downloads/K_OBP_XX_RI_01_05-10.pdf (Zugriff: 13.2.2014)
- Kindling, Maxi (2013): Qualitätssicherung im Umgang mit digitalen Forschungsdaten, in: Information - Wissenschaft & Praxis 64, H. 2-3, S. 137–148
- KIT-Archiv (2012): Startseite. URL: <http://www.archiv.kit.edu/> (Zugriff: 24.2.2014)
- KIT-Bibliothek (2014): Forschungsdatenmanagement am KIT. URL: <https://www.bibliothek.kit.edu/cms/forschungsdatenmanagement-am-kit.php> (Zugriff: 24.2.2014)
- Klump, Jens; Bertelmann, Roland (2013): D 8 Forschungsdaten, in: Kuhlen, Rainer; Semar, Wolfgang; Strauch, Dietmar (Hrsg.): Grundlagen der praktischen Information und Dokumentation. Berlin, Boston, De Gruyter Saur. DOI: 10.1515/9783110258264.575
- Kuberek, Monika (2012): Organisatorisch-technisches Konzept für eine Forschungsdaten-Infrastruktur in der TU Berlin. URL: http://www.sz-f.tu-berlin.de/fileadmin/f33_szf/TUB_Forschungsdaten_Konzept_lan_g_20120315.pdf (Zugriff: 11.3.2014)
- Lewis, Martin (2010): Libraries and the management of research data, in: McNight, S. (Hrsg.): Envisioning Future Academic Library Services. Initiatives, Ideas and Challenges. London, Facet Publishing, S. 145-168

- Liegmann, Hans; Neuroth, Heike (2010): Einführung, in: Neuroth, Heike; Oßwald, Achim (Hrsg.): Nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Göttingen, Nieders. Staats- u. Univ.-Bibl., Kap. 1:1-1:10
- Marcial, Laura Haak; Hemminger, Bradley M. (2010): Scientific data repositories on the Web. An initial survey., in: Journal of the American Society for Information Science and Technology 61, H. 10, S. 2029–2048. DOI: 10.1002/asi.21339
- Max-Planck-Gesellschaft (2003): Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. URL: http://openaccess.mpg.de/3515/Berliner_Erklaerung (Zugriff: 6.2.2014)
- Menne-Haritz, Angelika (2005): METS. Überblick und Anleitung. URL: http://www.loc.gov/standards/mets/METSOverview.v2_de.html (Zugriff: 3.2.2014)
- Mokrane, Mustapha (2013): Repository Audit and Certification Interest Group DSA. WDS Partnership Working Group. Case Statement. URL: <https://www.rd-alliance.org/filedepot?cid=115&fid=354> (Zugriff: 11.3.2014)
- Neuroth, Heike, Hrsg. (2012): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Boizenburg, Hülsbusch, URL: <http://nestor.sub.uni-goettingen.de/bestandsaufnahme/index.php> (Zugriff: 26.12.2013)
- Neuroth, Heike; Lossau, Norbert; Rapp, Andrea, Hrsg. (2013): Evolution der Informationsinfrastruktur. Forschung und Entwicklung als Kooperation von Bibliothek und Fachwissenschaft. Glückstadt, Werner Hülsbusch. URL: http://webdoc.sub.gwdg.de/univerlag/2013/Neuroth_Festschrift.pdf (Zugriff: 14.2.2014)
- Neuroth, Heike; Oßwald, Achim (2010): Nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.3. Göttingen, Nieders. Staats- u. Univ.-Bibl. URL: <http://nestor.sub.uni-goettingen.de/handbuch/index.php> (Zugriff: 01.12.2013)
- Organisation for Economic Cooperation and Development (OECD) (2007): OECD Principles and Guidelines for Access to Research Data from Public Funding. URL: <http://www.oecd.org/sti/sci-tech/38500813.pdf> (Zugriff: 30.12.2013)

- Pampel, Heinz; Bertelmann, Roland (2011): „Data Policies“ im Spannungsfeld zwischen Empfehlung und Verpflichtung, in: Büttner, Stephan (Hrsg.): Handbuch Forschungsdatenmanagement. Bad Honnef, Bock & Herchen, S. 49-61
- Pampel, Heinz; Vierkant, Paul; Scholze, Frank; Bertelmann, Roland; Kindling, Maxi; Klump, Jens; Goebelbecker, Hans-Jürgen; Gundlach, Jens; Schirmbacher, Peter; Dierolf, Uwe (2013): Making Research Data Repositories Visible. The re3data.org Registry, in: PLoS ONE 8, H. 11, e78080. DOI: 10.1371/journal.pone.0078080
- Piwowar, H. A. & Chapman, W. W. (2008): A review of journal policies for sharing research data, in: Proceedings of the 12th International Conference on Electronic Publishing. Toronto, Kanada 25.-27. Juni 2008.
URL: http://elpub.scix.net/data/works/att/001_elpub2008.content.pdf (Zugriff am 17.02.2011).
- PLOS One (o. J.): PLOS Editorial and Publishing Policies. Sharing of Materials, Methods, and Data. URL: <http://www.plosone.org/static/policies.action#sharing> (Zugriff: 7.2.2014)
- Razum, Matthias (2011): Systeme und Systemarchitekturen für das Datenmanagement, in: Büttner, Stephan (Hrsg.): Handbuch Forschungsdatenmanagement. Bad Honnef, Bock & Herchen, S. 123-138
- re3data.org team (2013): BioSharing and re3data.org cooperate on the collection and description of data repositories. URL: <http://www.re3data.org/2013/11/biosharing-and-re3data-cooperation/> (Zugriff: 8.1.2014)
- re3data.org team (2012): Memorandum of Understanding between DataCite and re3data.org. URL: <http://www.re3data.org/2012/04/memorandum-of-understanding-between-datacite-and-re3data-org/> (Zugriff: 8.1.2014)
- re3data.org team (2013): Memorandum of Understanding between OpenAIRE. URL: <http://www.re3data.org/2013/10/memorandum-of-understanding-between-openaire-and-re3data-org/> (Zugriff: 8.1.2014)
- Reed Elsevier (2010): Elsevier and PANGAEA take next step. URL: <http://www.reedelsevier.com/mediacentre/pressreleases/2010/Pages/elsevier-and-pangaea-take-next-step.aspx> (Zugriff: 7.2.2014)

- Riley, Jenn (2010): Seeing Standards.
URL: <http://www.dlib.indiana.edu/~jenlrile/metadatamap/>
(Zugriff: 3.2.2014)
- Rühle, Stefanie; Kompetenzzentrum Interoperable Metadaten (KIM) (o. J.): Kleines Handbuch Metadaten.
URL: <http://www.kim-forum.org/Subsites/kim/SharedDocs/Downloads/DE/Handbuch/metadaten.html;jsessionid=2260599284E03-DA94D30904D55B179D8.prod-worker3> (Zugriff: 3.2.2014)
- Rümpel, Stefanie (2011): Der Lebenszyklus von Forschungsdaten, in: Büttner, Stephan (Hrsg.): Handbuch Forschungsdatenmanagement. Bad Honnef, Bock & Herchen, S. 25-34
- Steinbuch Centre for Computing (SCC) (2013): Large-Scale Data Management & Analysis. URL: <http://www.scc.kit.edu/forschung/l sdf.php>
(Zugriff: 24.2.2014)
- Schaaf, Ines (2011): Forschungsdaten im Netz. Untersuchung von online verfügbaren Repositorien. Stuttgart, Hochschule der Medien, Bachelorarbeit
- Schmidt, Nora (2013): Forschungsdatenmanagement und Bibliothek. Verortung In Kooperationsnetzwerken, in: Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare 66, H. 2, S. 250-278.
URL: http://eprints.rclis.org/19579/1/vm_66_2013_2_Schmidt.pdf
(24.12.2013)
- Scholze, Frank; Müller, Uwe (2012): Aufbau und Vernetzung eines Repositoriumsm, in: Open-Access-Strategien für wissenschaftlichen Einrichtungen. Bausteine und Beispiele, S. 13-15. DOI: 10.2312/allianzoa.005
- Schulze, Matthias; Stockmann, Ralf (2013): Open Science und Networked Science: Offenheit und Vernetzung als Leitmotive und Visionen einer digitalen Wissenschaft im 21. Jahrhundert, in: Neuroth u. a. (Hrsg.): Evolution der Informationsinfrastruktur. Forschung und Entwicklung als Kooperation von Bibliothek und Fachwissenschaft. Glückstadt, Werner Hülsbusch, S. 31-38
- Schumann, Natascha (Hrsg.) (2012): Einführung in die digitale Langzeitarchivierung, in: Altenhöner, Reinhard (Hrsg.): Langzeitarchivierung von Forschungsdaten. Standards und disziplinspezifische Lösungen. Berlin, Scivero-Verl., S. 39-50

- Schwerpunktinitiative „Digitale Information“ der Allianz der Wissenschaftsorganisationen (2010): Grundsätze zum Umgang mit Forschungsdaten. URL: <http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsaeetze/> (Zugriff: 30.12.2013)
- Southall, Nathalie; Jisc (2013): Managing Research Data Programme 2011-13. URL: http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata.aspx (Zugriff: 7.1.2014)
- Spindler, Gerald; Hillegeist, Tobias (2010): Langzeitarchivierung wissenschaftlicher Primärdaten, in: Neuroth, Heike; Oßwald, Achim (Hrsg.): Nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Göttingen, Nieders. Staats- u. Univ.-Bibl., Kap. 16:1-16:24
- Spindler, Gerald; Hillegeist, Tobias (2011): Rechtliche Probleme der elektronischen Langzeitarchivierung von Forschungsdaten, in: Büttner, Stephan (Hrsg.): Handbuch Forschungsdatenmanagement. Bad Honnef, Bock & Herchen, S. 63-69
- Trusted Digital Repository (2010): European Framework for Audit and Certification of Digital Repositories. URL: <http://www.trusteddigitalrepository.eu/Site/Trusted%20Digital%20Repository.html> (Zugriff: 6.1.2014)
- Umlauf, Konrad; Lohnert, Peter (Hrsg.) (2011): Lexikon der Bibliotheks- und Informationswissenschaft. Bd. 1 A bis J. Stuttgart, Hiersemann
- Universität Bielefeld (2013): Resolution zu Forschungsdatenmanagement. URL: <http://data.uni-bielefeld.de/de/resolution> (Zugriff: 6.2.2014)
- Vierkant, Paul; Spier, Shaked; Rücknagel, Jessika; Pampel, Heinz; Gundlach, Jens; Fichtmüller, David; Kindling, Maxi; Kirchhoff, Agnes; Göbelbecker, Hans-Jürgen; Klump, Jens; Kloska, Gabriele; Reuter, Evelyn; Semrau, Angelika; Schnepf, Edeltraud; Skarupianski, Michael; Bertelmann, Roland; Schirnbacher, Peter; Scholze, Frank; Kramer, Claudia (2013): Schema for the description of research data repositories. DOI: 10.2312/re3.004
- Weichselgartner, Erich; Günther, Armin; Dehnhard, Ina (2011): Archivierung von Forschungsdaten, in: Büttner, Stephan (Hrsg.): Handbuch Forschungsdatenmanagement. Bad Honnef, Bock & Herchen, S. 191-202

- Wellcome Trust (o. J.): Statement on genome data release.
URL: <http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTD002751.htm> (Zugriff: 6.2.2014)
- Wilson, James (2011): Welcome to the DaMaRO website. URL: <http://damaro.oucs.ox.ac.uk/> (Zugriff: 27.2.2014)
- Wilson, James; Jeffreys, Paul (2013): Towards a Unified University Infrastructure. The Data Management Roll-Out at the University of Oxford, in: International Journal of Digital Curation 8, H. 2, S. 235-246. DOI: 10.2218/ijdc.v8i2.287
- Winkler-Nees, Stefan (2010): Forschungsdaten besser nutzen. Strategien und Ansätze der DFG. Fachreferententagung Soziologie, 26.11.2010 in Berlin. URL: http://staatsbibliothek-berlin.de/fileadmin/user_upload/zentrale_Seiten/faecher/pdf/Winkler_Nees_VDB_26NOV2011.pdf (Zugriff: 18.9.2013)
- Zentrum für Angewandte Kulturwissenschaft (ZAK) (2013): Center of Digital Tradition (ZAK-CODIGT). URL: <http://www.zak.kit.edu/codigt.php> (Zugriff: 24.2.2014)
- Zentrum für Angewandte Kulturwissenschaft und Studium Generale-Center of Digital Tradition (ZAK-CODIGT) (2012): dLZA von Forschungsdaten am KIT. Auswertungsdiagramme
- Zentrum für Angewandte Kulturwissenschaft und Studium Generale-Center of Digital Tradition (ZAK-CODIGT) (2012): dLZA von Forschungsdaten am KIT. Auswertungssynopse