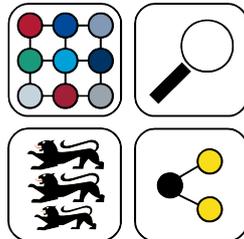
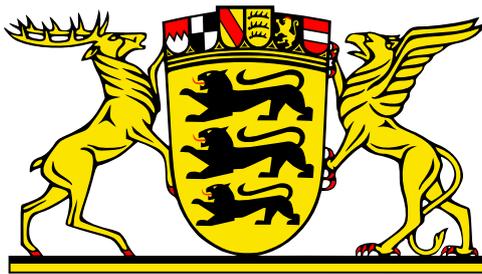


Öffentlicher Abschlussbericht von bwFDM-Communities



Wissenschaftliches Datenmanagement an den
Universitäten Baden-Württembergs



Gefördert durch das Ministerium für Wissenschaft,
Forschung und Kunst Baden-Württemberg

Projektzeitraum - 1. 1. 2014 bis 30. 9. 2015

Autoren

Projektleitung

Frank Tristram (KIT)
Kontakt: Frank.Tristram@kit.edu

Key-Accounter

Peter Bamberger (Universität Ulm)
Uğur Çayoğlu (KIT)
Dr. Jörg Hertzner (Universität Stuttgart)
Johannes Knopp (Universität Mannheim)
Jonas Kratzke (Universität Heidelberg)
Jessica Rex (Universität Konstanz)
Fabian Schwabe (Universität Tübingen)
Dr. Denis Shcherbakov (Universität Hohenheim)
Dr. Dieta-Frauke Svoboda (Universität Tübingen)
Dennis Wehrle (Universität Freiburg)

Projektverantwortung

Prof. Dr. Achim Streit (KIT)
Kontakt: Achim.Streit@kit.edu

Inhaltsverzeichnis

1	Einleitung	1
1.1	Projektumfang und Ziele	2
1.2	Aufbau des Berichtes	2
1.3	Methoden	3
2	Strukturelle Rahmenbedingungen	8
2.1	Urheberrecht und Datenschutz	8
2.2	Informationsangebote	15
2.3	Wissenschaftliche Datenkultur	26
3	Digitale Forschungsdatenverarbeitung	37
3.1	Forschungsnahе Unterstützung	37
3.2	Wissenschaftliche Softwaretools	41
3.3	Formate und einheitliche Standards	46
4	Datenerhebung und -austausch	53
4.1	Daten außeruniversitärer Einrichtungen	53
4.2	Erwartungen an die Digitalisierung	58
4.3	Wissenschaftliche Zusammenarbeit	64
5	Informationsbewahrung	74
5.1	Projekt- und Datendokumentation	74
5.2	Repositorien für die Publikation von Forschungsdaten	87
5.3	Archivierung	94
6	Weitere Aspekte	110
6.1	IT-Infrastruktur	110
6.2	IT-Support	120
6.3	Lizenzierung	123
6.4	Finanzierung	126
6.5	Open Science	128
6.6	Vorbehalte gegenüber dem Forschungsdatenmanagement	135
7	Fazit	138
	Anhang	144

1 Einleitung

Schon Isaac Newton empfand seine Errungenschaften nur als einzelnes Glied in einer langen wissenschaftlichen Wertschöpfungskette.¹ Damit diese Kette funktioniert, müssen bekannte Forschungsergebnisse zur richtigen Zeit, in der richtigen Form bei nachfolgenden Forschenden ankommen. Das zu erreichen, ist Ziel von Forschungsdatenmanagement (FDM). So betrachtet, ist FDM in der Wissenschaft gut vergleichbar mit Supply-Chain-Management (SCM) in der Wirtschaft, welches Wertschöpfungsketten über mehrere Akteure hinweg optimiert und dafür sorgt, dass Rohstoffe in der Verpackung und Form ankommen, die am besten verwendbar sind. Firmenübergreifendes SCM ist durch immer komplexere Wertschöpfungsketten zu einem wichtigen Erfolgsfaktor in vielen modernen Produktionsstätten geworden, die dafür große Abteilungen unterhalten. Beim FDM sind nicht Unternehmen, sondern Forschungsgruppen die Einzelakteure der Wertschöpfungskette. Als Besonderheit in der Wissenschaft kommt hinzu, dass die „Supply-Chain“ ein Zyklus ist und Forschungsergebnisse wieder als Rohstoff für weitere Forschung gedacht sind. In der Wissenschaft steht übergeordnetes FDM noch am Anfang, ist aber aufgrund immer komplexerer Forschungsprozesse – wie wir zeigen werden – schon jetzt ein wichtiger Effektivitätsfaktor. Die Bedeutung der dargestellten Prozesskettenperspektive ist den forschungsrelevanten Institutionen in Deutschland inzwischen bewusst. Baden-Württemberg bettet FDM darum in seine E-Science-Strategie² ein. Diese zielt darauf ab, Forschenden zu ermöglichen, Daten barrierefrei aufzubereiten, zu verarbeiten sowie auffindbar und vor allem nachnutzbar abzulegen.

Die Art und Weise wie wir mit Forschungsdaten umgehen, wird stark beeinflussen, wie schnell unser Wissen und damit auch unser Wohlstand wachsen wird. Durch einen offeneren Umgang mit Daten in der Wissenschaft, den öffentlichen Institutionen und der Industrie in Baden-Württemberg könnten jährlich schätzungsweise Mittel im zwei- bis dreistelligen Millionenbereich effizienter eingesetzt werden, die bisher in Datensuche, Doppelarbeit und, mangels besserer Möglichkeiten, in ineffizienten Arbeitsmethoden „verschwinden“. Der Mehrwert einer sichtbaren Datenarchivierung und der Möglichkeit einer sekundären Nutzung besteht darin, das Potential bereits vorhandenen Wissens voll auszuschöpfen. Gebraucht wird dafür eine grundlegende Vereinfachung der Methoden der Datengenerierung und -aufbereitung sowie der Verwendung und Zugänglichmachung von Daten, sodass ein breites Spektrum von Akteuren mit unterschiedlichen Methoden unmittelbar auf vorhandene Fakten aufbauen kann.

¹ „If I have seen further it is by standing on y^e shoulders of Giants.“; Newton zu Robert Hooke, 5 Feb. 1676, <http://www.isaacnewton.org.uk/essays/Giants> [letzter Zugriff 15.3.2016].

² E-Science – Wissenschaft unter neuen Rahmenbedingungen. Fachkonzept zur Weiterentwicklung der wissenschaftlichen Infrastruktur in Baden-Württemberg, https://mwk.baden-wuerttemberg.de/fileadmin/redaktion/m-mwk/intern/dateien/pdf/Forschung/066_PM_Anlage_E-Science_Web.pdf, 89 [letzter Zugriff 30.6.2015].

1.1 Projektumfang und Ziele

Im Rahmen von *bwFDM-Communities* konnten für die Universitäten in Baden-Württemberg die Bedarfe der Forschenden im Bereich des FDMs erhoben und ausgewertet werden. Zukünftige Maßnahmen können auf diesen Erkenntnissen aufbauen, um den Umgang mit Forschungsdaten zu erleichtern. Das Projekt *bwFDM-Communities* hat von Januar 2014 bis Juni 2015 an allen baden-württembergischen Universitäten Kontakt zu dem wissenschaftlichen Personal gesucht und deren Bedarfe und Wünsche ausgewertet. Dabei wurden in 627 Interviews 779 Einzelpersonen zur Verwendung von Daten in ihrer Forschung befragt. Im Vordergrund standen dabei Wünsche und Mittel zur besseren Ausschöpfung eines in Daten liegenden Forschungspotentials. Diese Studie im Auftrag des Ministeriums für Wissenschaft, Forschung und Kunst Baden-Württemberg (MWK) soll beleuchten, was für die Forschenden die wichtigen Meilensteine beim digitalen Wandel sind, um die Ziele der E-Science-Strategie zu erreichen und konkrete Umsetzungsschritte für diesen Weg vorzuschlagen.

1.2 Aufbau des Berichtes

1.2.1 Wie Sie den Bericht lesen können

Alle Kapitel sind unabhängig voneinander lesbar und bei Beziehungen untereinander mit Querverweisen ausgestattet. Neben herausgegriffenen Themen ist es auch möglich einzeln über die Bedarfe oder die Handlungsempfehlungen zu lesen. Welche FDM-Inhalte besonders häufig von den Forschenden angesprochen wurden, steht in aller Kürze am Ende des Berichts (7. Fazit). Der für die Befragungen benutzte Interviewleitfaden ist dem Anhang des Berichtes beigelegt. Ebenso findet sich dort ein Abkürzungsverzeichnis aller verwendeten Fachbegriffe. Bei Interesse an ganz bestimmten Themen empfiehlt sich nach dem Lesen der entsprechenden Kapitel auch ein Blick auf die Details der zugrundeliegenden [Daten](#)³, da aufgrund der Fülle an Information, viele Bemerkungen der Forschenden nur in abstrahierter Form im Bericht angeführt werden konnten. Für Projektideen finden sich dort wertvolle Hinweise zur konkreten Ausgestaltung und Umsetzung.

1.2.2 Erklärung der Berichtstruktur

In dem vorliegenden Abschlussbericht sind zum einen Themenbereiche beschrieben, die eng mit FDM zusammenhängen (Berichtsteile 2–5), zum anderen auch mehrere Abschnitte, die zwar für Forschungsdaten relevant sind, aber mit deren Management nur indirekt zu tun haben. Diese Inhalte können als nützlicher Beifang der Interviews betrachtet werden und sind in 6. [Weitere Aspekte](#) zusammengefasst. Um die Länge des Fazits nicht zu überspannen wurden diese Themen im Fazit nicht berücksichtigt.

Die Bereiche [2. Strukturelle Rahmenbedingungen](#), [3. Digitale Forschungsdatenverarbeitung](#), [4. Datenerhebung und -austausch](#) und [5. Informationsbewahrung](#) haben einen unmittelbaren FDM-Bezug. Zu den darin enthaltenen Kapiteln wie [2.2. Informationsangebote](#) oder

³ Datenseite von *bwFDM-Communities*: <https://bwfdm.scc.kit.edu/daten> [letzter Zugriff 29.2.2016].

5.3. **Archivierung**, ist in Klammern hinter ihrer Überschrift die Anzahl der zugrundeliegenden User-Storys⁴ angegeben. Diese Kapitel haben eine einheitliche Gliederung in Unterabschnitte:

- Einleitung
- Bedarf der Wissenschaftsgemeinschaft
- Status Quo
- (E-Science-Bestrebungen)
- Gegenwärtige Handlungsempfehlungen

1.2.2.1 Details zu den Inhalten der Kapitelabschnitte

In jeder Einleitung sind die jeweils zugehörigen User-Storys nach Wissenschaftsbereichen aufgeschlüsselt. Daran schließt sich die Auswertung der User-Storys im Abschnitt „Bedarf der Wissenschaftsgemeinschaft“ an. Hier erfolgt die Kategorisierung der User-Storys in Unterthemen, die mit einer Meta-User-Story wiedergegeben werden. Eine Meta-User-Story bündelt mehrere User-Storys zu einer etwas abstrahierten Aussage für eine Gruppe von Forschenden mit ähnlichen Belangen. Die in der jeweiligen Überschrift in Klammern angegebene Zahl zeigt abermals die Anzahl der im Unterthema behandelten User-Storys.

Im Abschnitt „Status quo“ wird den Wünschen der Forschenden der derzeitige Stand von Infrastruktur und Anwendungspraxis bzw. aktueller Methodik gegenübergestellt. Für einige Themenbereiche gibt es einen ergänzenden Abschnitt „E-Science-Bestrebungen“, in dem der Ist-Zustand um die Empfehlungen des E-Science-Fachkonzepts (ESFK) zur Weiterentwicklung der wissenschaftlichen Infrastruktur erweitert wird. Jeder Themenbereich schließt mit einem Abschnitt „Gegenwärtige Handlungsempfehlungen“. Dort finden sich Vorschläge, wie man den zuvor als wichtig identifizierten Wünschen entsprechen kann und weitere Erkenntnisse, die sich aus der Analyse unserer Daten ergeben haben. Die detaillierten Beschreibungen der einzelnen Themenbereiche mit ihren Unterthemen werden in 7. Fazit auf die wichtigsten Kernerkenntnisse und Handlungsempfehlungen reduziert.

1.3 Methoden

Der folgende Methodenbericht enthält einen Überblick über die Studienmethodik, mit der die später vorgestellten Ergebnisse des Projektes *bwFDM-Communities* erhoben, ausgewertet und aufbereitet wurden.

Bei der durchgeführten Studie handelt es sich um eine qualitative, nicht-repräsentative Erhebung, die mithilfe von halbstrukturierten Face-to-Face Interviews durchgeführt wurde. An jeder der Universitäten Freiburg, Heidelberg, Hohenheim, Konstanz, Mannheim, Stuttgart, Tübingen, Ulm und des KIT war für die Durchführung und Auswertung der Interviews ein Key-Accounter verantwortlich. In Freiburg und Tübingen, teilten sich jeweils zwei Key-Accounter die Verantwortung. Die Projektleitung war am KIT angesiedelt.

⁴ Das User-Story-Konzept wird in 1.3.2. Auswertung und User-Storys erklärt.

1.3.1 Datenerhebung durch Interviews

Die Feldzeit der Datenerhebung erstreckte sich von Januar 2014 bis November 2014 und war in drei Phasen unterteilt. Die erste Phase erstreckte sich vom 1.1.2014 bis zum 30.4.2014.

In dieser sogenannten Kontaktphase legten die Key-Accounter für ihren jeweiligen Standort eine Kontaktliste aller wissenschaftlichen Communitys an und identifizierten einen geeigneten Ansprechpartner in jeder dieser Communitys (soweit möglich einen pro Arbeitsgruppe). Anschließend wurden diese Ansprechpartner von den Key-Accountern kontaktiert. An den meisten Standorten wurde eine Kombination aus E-Mails und Telefonaten gewählt. In einigen Fällen wurde auf schriftliche Einladungen oder persönlichen Kontakt gesetzt.

In der Testphase, die sich vom 1.5. bis 28.5.2014 erstreckte, wurden die ersten Ansprechpartner mithilfe eines ersten Gesprächsleitfadens befragt. Dieser Leitfaden wurde auf Basis der in diesen Testinterviews gewonnenen Erkenntnisse anschließend überarbeitet. Hierbei wurden vor allem Fragen sinnvoller gruppiert, umformuliert und einige Fragen zusammengefasst bzw. gestrichen. In der Testphase wurden 67 von insgesamt 627 Interviews geführt, was einer üblichen Test-Quote von ca. 10 % entspricht.

In der Bedarfsermittlungsphase wurde vom 29.5. bis zum 31.10.2015 der Hauptteil der Interviews (560) geführt. Da einige Interviews mit mehr als nur einem Gesprächspartner erfolgten, ergibt sich insgesamt eine Anzahl von 779 interviewten Einzelpersonen. Die Systematisierung und Auswertung der Daten erfolgte in der Präzisionsphase und in der Abschlussphase (1.11.2014–31.3.2015 und 1.4.–30.6.2015).

Auswahl der Interviewpartner und Ausschöpfungsquote Alle an den Landesuniversitäten vorhandenen wissenschaftlichen Gruppen galten als potentielle Gesprächsinteressenten und alle Interviewten als Repräsentierende einer „Community“. Eine wissenschaftliche Community ist hierbei definiert als eine Gruppe von Forschenden, die an einem Standort gemeinsam an einer eng verwandten wissenschaftlichen Thematik arbeitet. In der Regel hat die Gruppenleitung bestimmt, wer die Community vor uns repräsentieren sollte.⁵ Dieses Vorgehen sorgte weder in der Hierarchie noch im Fachgebiet für repräsentative Verteilungen. Durch Angabe der DFG-Fachbereiche haben wir Raum für Entzerrung und Übertragung auf andere Einrichtungen gelassen. Neben Führungskräften aus Instituten und Forschungsgruppen haben wir auch Postdocs, IT-Kräfte, Promovierende und in einigen Fällen auch (Master- und Bachelor-)Studierende sowie Gastwissenschaftler befragt, sodass ein breites Meinungsbild durch alle wissenschaftlichen Hierarchieebenen eingefangen werden konnte. Es wurde zudem versucht darauf zu achten, mit allen relevanten Sektionen und Fakultäten jedes Standortes Gespräche zu führen, um Ergebnisse aus einer möglichst breit gestreuten Grundgesamtheit zu erhalten.

Die Ausschöpfungsquote war aufgrund eines eigenen Key-Accounters in jedem Standort gut. Sie beträgt je nach Rechenweise 31 - 38 %. Die zwei Werte kommen daher, dass manchmal mehrere Personen gleichzeitig interviewt wurden (die zum Teil gar nicht aufgefordert worden waren). Diese nicht mitzuzählen führt zu der niedrigeren Quote. Die Quote schwankt aber stark zwischen den Standorten. Großskalige Ansätze mit bis zu 750 per E-Mail angespro-

⁵ Das war regelmäßig auch direkt sie selbst.

chenen Personen in 692 Arbeitsgruppen (KIT) erlauben bei angestrebten 100 Interviews pro Standort naturgemäß nur eine schlechtere Ausschöpfung (erreicht wurden dort etwa 20 %) und drücken den Schnitt enorm. Werden die einzelnen Standortquoten gemittelt ergibt sich eine Ausschöpfungsquote von 44 - 60 % (je nach Rechenweise, s.o.).

Interviewleitfaden Der Interviewleitfaden, der den im Projekt geführten Gesprächen zugrunde liegt, kann im Anhang eingesehen werden. Er besteht aus sechs Abschnitten (Einleitende Fragen, Umfeld der Daten, Prozess der Datenverarbeitung, Lagerung, Publikation und Data Sharing, Allgemeine Angaben), die sich am Lebenszyklus von Forschungsdaten orientieren.

Der Leitfaden enthält 27 Hauptfragen und 13 Unterfragen. Davon sind 27 Fragen offen formuliert, zehn halboffen und drei sind geschlossene Fragen. Zudem enthält der Leitfaden neun einfache Filterfragen (1d, 9a, 11a, 12a, 14, 17a, 19b, 20a und 22a). Die Anzahl der Antworten schwankt erheblich, da einerseits Mehrfachantworten erlaubt waren und es den befragten Personen jederzeit freistand, eine Frage nicht zu beantworten.

Zusätzlich zu den eigentlichen Fragen erhebt der Leitfaden Angaben zu Fachbereich, Position der befragten Person in der Community sowie die grobe Größe der repräsentierten Arbeitsgruppe. Jedes Interviewdokument enthält weiterhin eine vom Key-Accounter für den Standort einzigartige Identifikationsnummer.

Die Art der Erfassung der Gesprächsinhalte blieb jedem Key-Accounter selbst überlassen. So kamen sowohl handschriftliche Notizen als auch ein Transkript der Äußerungen, ein Fragebogentool und Tabellen – während oder auch nach dem Interview – zum Einsatz. Es wurde aber ein standardisiertes CSV-Format festgelegt, in dem die Interviewdaten mit denen anderer Standorte zusammengeführt werden konnten.

Interviewpersonal und Interviewdauer Die Interviews wurden an den Universitäten in der Regel von den Key-Accountern selbst durchgeführt. Die Interviewdauer lag im Schnitt bei circa 75–80 Minuten, wobei die Varianz, dem offenen Gesprächscharakter der Interviews geschuldet, entsprechend groß ausfiel. So dauerte das kürzeste Interview lediglich 15 Minuten, während das längste Gespräch 260 Minuten in Anspruch nahm.

1.3.2 Auswertung und User-Storys

Die in den Gesprächen erfassten, explizit geäußerten Wünsche der befragten Personen wurden notiert und in eine einheitlichen Struktur übertragen: „In der Rolle als X wünsche ich mir Y, um Z zu können“. Diese strukturierten Bedarfsäußerungen werden als User-Storys bezeichnet. Die Anzahl der notierten User-Storys schwankte pro Interview je nach Gesprächsbedarf und Persönlichkeit der Befragten und der Interviewer stark. Im Schnitt entfielen auf ein Interview zwischen drei und vier User-Storys, aber auch Ausreißer von 0 bis fast 20 hat es gegeben.

Diese User-Storys wurden auf einer Online-Plattform zusammengetragen und von allen Mitarbeitern einem Reviewprozess unterzogen, nach dessen Ende die User-Storys inhaltlich verständlich und formell richtig und vollständig sein sollten. Anschließend wurden die User-Storys zu verschiedenen Themen gruppiert. Die gebildeten Kategorien entsprechen im Großen

und Ganzen den thematischen Kapiteln des vorliegenden Abschlussberichts. Die User-Stories wurden anhand der Fachsystematik der Deutschen Forschungsgesellschaft (DFG)⁶ zugeordnet. Der im Bericht häufig benutzte Begriff „Wissenschaftsbereich“ wird entsprechend dieser Fachsystematik verwendet (die vier DFG-Wissenschaftsbereiche finden sich beispielsweise in Abb. 1).

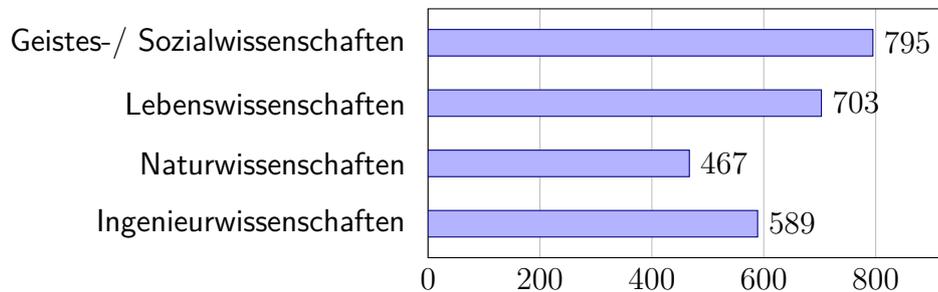


Abb. 1: Verteilung der User-Stories nach der Fachsystematik der DFG

Interpretation der Zahlen des Berichts Die Ergebnisse der Auswertungen der Interviewtexte und User-Stories finden sich im vorliegenden Abschlussbericht sowohl als Grafiken und Tabellen als auch im Fließtext als Zahlen- bzw. Prozentangaben wieder. Zum besseren Verständnis dieser Angaben muss beachtet werden, dass Antworten auf offene Fragen sowie Wünsche spontan von der befragten Person formuliert wurden und daher schon bereits Inzidenzen von $n = 5$ (oder mehr) als signifikant gelten können. Stets vorhandene Umwelteinflüsse auf die Befragten, bspw. ein tagesaktuelles Forschungsproblem oder der Ausfall eines Systems in der vorangegangenen Woche, bestimmen erheblich was spontan zu einem Thema geantwortet wird. Geben bspw. fünf von 500 befragten Personen eine bestimmte Antwort oder äußern einen Wunsch, so ist daraus keinesfalls abzulesen, dass die restlichen 495 Personen diese Antwort oder diesen Wunsch ablehnten. Vielmehr ist das gehäufte Auftreten einer spontan gegebenen Antwort als Indiz für die Abbildung einer signifikant häufig auftretenden Realität zu werten. Auf der anderen Seite hat der Interviewleitfaden durch die Richtungsvorgabe des Gesprächs natürlich auch gewisse Äußerungen verursacht. Weiterhin kann es bei den Ergebnissen aus halboffenen und geschlossenen Fragen dazu kommen, dass die Gesamtzahl der Antworten die Gesamtzahl der befragten Personen übersteigt, wenn Mehrfachantworten zulässig waren. Der vielleicht wichtigste Punkt ist, dass unsere Zahlen, trotz aller Systematik, Gesprächen entstammen und in der Regel nur qualitative Aussagen zulassen, denn die Auswahl der Interviewenden und erhaltenen Aussagen ist grundsätzlich nicht verzerrungsfrei. So sind beispielsweise unmotivierte oder ohnehin zeitlich stark beanspruchte Personen vermutlich weniger gesprächsfreudig. Außerdem variiert die Größe der Forschungsgruppen. Die Regel, mit jeder Arbeitsgruppe einmal zu sprechen, überrepräsentiert systematisch kleine Gruppen, was aber durchaus auch dem Ansinnen dieses Projekts entgegenkommt auch den kleinen Gruppen ein Gehör zu verschaffen. In diesem Sinne sind viele Zahlen vor allem relativ,

⁶ http://www.dfg.de/download/pdf/dfg_im_profil/gremien/fachkollegien/amtperiode_2012_2015/fachsystematik_2012_2015_de_grafik.pdf - [letzter Zugriff: 30.6.2015].

und sollten auf die Grundgesamtheit bezogen oder relativ zu anderen Verteilungen im Bericht betrachtet werden.

1.3.3 Anonymisierung und Datenschutz

Alle Befragten haben freiwillig an der Erhebung teilgenommen und waren über ihre Freiwilligkeit, den Zweck sowie die projektgebundene Datenverarbeitungsdauer informiert. Jedem Gesprächspartner wurde, auf Wunsch schriftlich, versichert und erklärt, dass sichergestellt wird, dass die erhobenen Daten nach den Standards des Bundes- und Landesdatenschutzgesetzes behandelt werden. So wurden die Listen der Gesprächspartner an den jeweiligen Standorten verschlüsselt und separat von den Interviewinhalten aufbewahrt. Eine Zusammenführung kann nur durch den jeweiligen Key-Accounter oder, in begründeten Fällen (z.B. längere Krankheit eines Mitarbeiters) entsprechend delegiert, durch die Projektleitung erfolgen. Dritte haben dazu keinen Zugang. Nach Beendigung der Datenauswertung werden die Daten der Gesprächspartner gelöscht. Zusätzlich erfolgt die Auswertung der Gesprächsinhalte zusammengefasst in größeren Gruppen. Wo Auszüge und / oder Zitate aus einzelnen Gesprächen publik gemacht werden, wird sichergestellt, dass zuvor denkbare Identifikationsmerkmale entfernt wurden.

2 Strukturelle Rahmenbedingungen

2.1 Urheberrecht und Datenschutz (117)

2.1.1 Einleitung

Urheber- und datenschutzrechtliche Aspekte können in sämtlichen Phasen des Forschungsprozesses eine Rolle spielen. Forschende sind keine Juristen und brauchen hier in der Regel Handreichungen und Beratung, um in einer vernetzten Welt allen Anforderungen des Urheberrechts und des Datenschutzes gerecht zu werden. Es erscheint plausibel, dass Forschende vor allem während der Publikationsphase mit Fragen aus diesen Rechtsbereichen in Kontakt kommen. Das kann auch durch eine nähere Betrachtung der User-Stories bestätigt werden, wobei sich Unterschiede zwischen den Disziplinen ergeben.

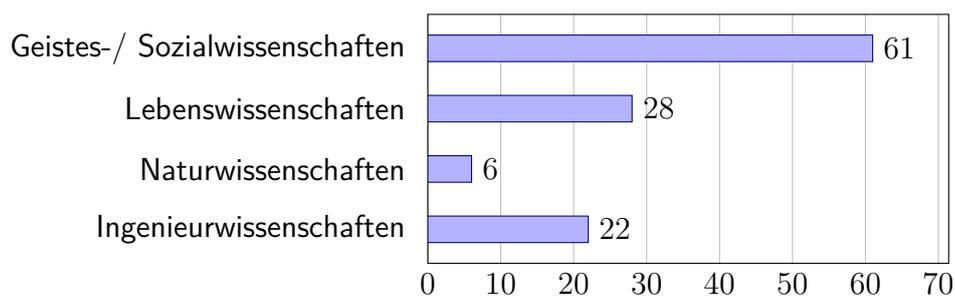


Abb. 2: Verteilung der User-Stories nach der Fachsystematik der DFG

Eine Übersicht über die Relevanz für die unterschiedlichen Wissenschaftsbereiche und der einzelnen Themen findet sich in [Abb. 2](#). Wie dort zu sehen ist, sind vor allem die geistes- und sozialwissenschaftlichen Gebiete sehr präsent, was auf die Benutzung von personenbezogenem Forschungsmaterial in diesen Bereichen zurückzuführen ist. Auch die Lebenswissenschaften haben sich im medizinischen Bereich gehäuft mit Datenschutz zu befassen. Die urheberrechtlichen Bedarfe der Forschenden erstrecken sich über alle Wissenschaftsbereiche, sind aber vor allem in den Ingenieurwissenschaften sehr ausgeprägt, wohingegen rechtliche Themen für die Naturwissenschaften mit Abstand die geringste Rolle spielen.

2.1.2 Bedarf der Wissenschaftsgemeinschaft

Um herauszufinden, welche Bedeutung urheber- und/ oder datenschutzrechtliche Aspekte beim FDM haben, wurden folgende zwei Fragen im Interviewleitfaden angesprochen:

1. „Forschen Sie an personenbezogenen Daten?“

2. „Welche rechtlichen Ansprüche haben Dritte an Ihren Daten?“

Etwa ein Viertel der im Rahmen des Projektes befragten Forschenden gaben hier an, an personenbezogenen Daten zu forschen. Knapp 40 % besitzen nach eigenen Angaben kein exklusives Recht an all ihren Daten.

Von den insgesamt 2554 User-Stories setzen sich 117 explizit mit urheber- und / oder datenschutzrechtlichen Fragestellungen (Problemen) auseinander. Der Inhalt der User-Stories dieses Kapitels kann in vier Themen gegliedert werden, die in den nachfolgenden Abschnitten näher dargestellt sind.

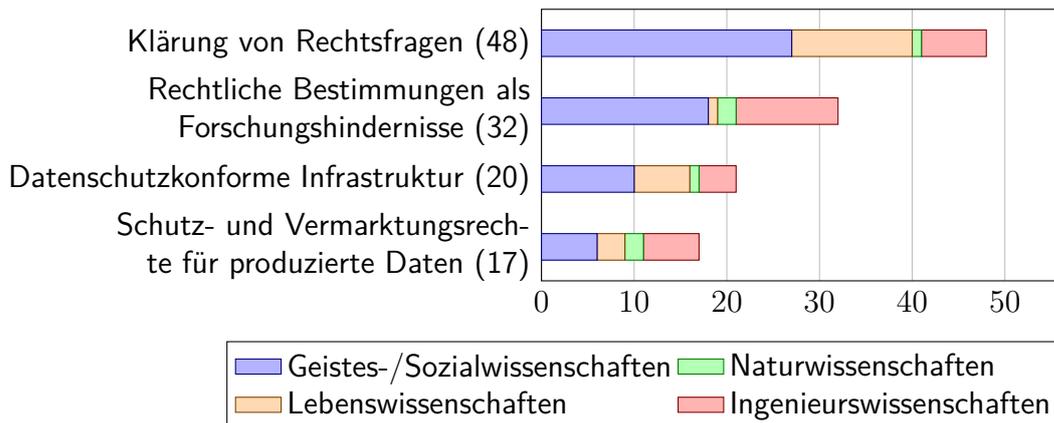


Abb. 3: Verteilung rechtlicher Themen nach der Fachsystematik der DFG

Abb. 3 zeigt wie relevant die Unterthemen für einzelne Fachbereiche sind. Während die ersten beiden Themen sich sowohl mit Datenschutz, als auch mit Urheberrecht befassen, sind die übrigen Themen nur je einem der Rechtsfelder gewidmet. Es ist erkennbar, dass die Datenschutzbestimmungen in den Lebenswissenschaften in sofern akzeptiert zu sein scheinen, als dass diese in der Regel nicht als Hindernisse wahrgenommen werden, sondern in der Hauptsache einen „Klärungsbedarf“ erzeugen.

2.1.2.1 Klärung von Rechtsfragen (48)

„Als Forschende, wünschen wir uns rechtliche Hilfsmittel und Beratungsangebote, um unsere Fragen zu den Themen Urheberrecht und Datenschutz zu klären.“

Mit 48 User-Stories handelt es sich bei Wünschen zur Klärung von Rechtsfragen um einen großen Bedarf. In den User-Stories werden primär Wünsche zu Informations- und Beratungsangeboten geäußert. Einige Rechtsfragen sollen aber auch von Ethikräten, durch Zertifikate oder ähnliches geklärt werden, damit sich bei den Forschenden ein begründbares Gefühl von Rechtssicherheit einstellen kann. Durch wen eine Beratung erfolgen soll, geht aus keiner der User-Stories explizit hervor.

2.1.2.2 Rechtliche Bestimmungen als Forschungshindernisse (32)

„Als Forschende, wünschen wir uns flexiblere Bestimmungen im Urheberrecht und Datenschutz, um die Freiheit der Forschung sicherzustellen.“

In den 32 User-Stories wird herausgearbeitet, wie die Forschenden einen Teil der geltenden Bestimmungen des Urheberrechtsgesetz und (Landes-)Datenschutzgesetzes als hinderlich bzw. einschränkend empfinden. In Bezug auf das Urheberrecht fokussiert sich die Kritik vor allem auf die aktuelle Ausgestaltung der Schranken im Urheberrechtsgesetz (UrhG). Die Orientierung an einzelnen Ausnahmetatbeständen, beispielsweise für die Verwendung von lizenzgeschützten Materialien in Forschung und Lehre, bei der Nachnutzung von Forschungsdaten und bei dem Bezug digitaler Kopien, wird als hinderlich angesehen. Gefordert wird eine Revision der Schrankenregelung im UrhG. Als Referenzmodell wünschen sich manche Forschenden die in den USA angewandte Fair-Use-Regelung.

In Bezug auf das (Landes-)Datenschutzgesetz lassen sich die Kritikpunkte so zusammenfassen, dass es zum einen die Zugänglichkeit mancher Daten unterbindet und zum anderen, dass die in anonymisierter Form zugänglichen Daten für manche Forschungszwecke ungeeignet sind. Das zweite Argument wird sowohl von Datenproduzenten als auch von Datenkonsumenten vorgebracht.

2.1.2.3 Datenschutzkonforme IT- Infrastruktur (20)

„Als Forschende, wünschen wir uns zentrale IT-Infrastruktur, um mit personenbezogenen Daten zu arbeiten.“

In diesem Unterthema wird der Bedarf an IT- Infrastruktur, die sämtlichen datenschutzrechtlichen Anforderungen gerecht wird, zum Ausdruck gebracht. Dieser Bedarf bezieht sich sowohl auf den Austausch, als auch auf die Speicherung und (Langzeit-) Archivierung von Daten. Die Forschenden sind sich oft der rechtlichen Komplikationen einer solchen Infrastruktur nicht bewusst. Ein ähnliches Problem ist in [4.3. Wissenschaftliche Zusammenarbeit](#) zu beobachten. Wir gehen im Status quo und in den Handlungsempfehlungen auf diese Probleme ein, sehen aber das rechtliche Problem gegenüber dem technischen, zumindest bei unsensiblen Daten, im Vordergrund.

2.1.2.4 Schutz- und Vermarktungsrechte für produzierte Daten (17)

„Als Forschende mit selbst erzeugten Daten, wünschen wir uns Unterstützung bei Schutz- und Vermarktungsrechten, um über unsere Daten bestimmen zu können.“

Anders als in [2.1.2.2. Rechtliche Bestimmungen als Forschungshindernisse](#) wird in diesen User-Stories aus Sicht des Datenproduzenten bzw. Urhebers argumentiert. Das Hauptanliegen dieser User-Stories ist es, die Schutz- und Vermarktungsrechte der Forschenden gegenüber Externen (d.h. Verlage, Datenbankbetreiber, Kooperationspartner aus der Industrie und potentielle Nachnutzer) zu stärken und darauf hinzuweisen, dass für sensible Daten und bestimmte Forschungsk Kooperationen, bezüglich Forderungen nach mehr Open Science, Spielraum für Ausnahmen existieren muss.

2.1.3 Status quo

Im Folgenden wird der Status quo in Baden-Württemberg jeweils in Bezug auf die vorab skizzierten Bedarfe erläutert.

2.1.3.1 Informations- und Beratungsangebote

Es gibt an allen neun Landesuniversitäten Ansprechpartner für urheber- und /oder datenschutzrechtlichen Fragestellungen.

Bei der Beratung zu datenschutzrechtlichen Fragestellungen kooperieren alle Standorte außerdem mit der zentralen Datenschutzstelle der baden-württembergischen Universitäten (ZENDAS). Die Beratung erfolgt in einem First-Level-Support an den Universitäten.⁷ Von dort wird sich bei Bedarf mit ZENDAS in Verbindung gesetzt. Lediglich Mitarbeiter der Universität Stuttgart können sich direkt an ZENDAS bzw. den Leiter von ZENDAS wenden. Dieser ist gleichzeitig behördlich bestellter Datenschutzbeauftragter der Universität Stuttgart. Eine weitere Sonderstellung nimmt das KIT ein, welches als Landes- und Bundesbehörde ebenfalls über eine eigene behördliche Datenschutzbeauftragte verfügt.

Zum Datenschutz können sich Forschende auf der Homepage der ZENDAS⁸ umfassend informieren. Dort ist sehr viel Material gesammelt, doch der Fokus der ZENDAS ist wesentlich breiter als FDM, was zunächst abschreckend wirken könnte, denn die ZENDAS behandelt auch viele Themen zu Verwaltungsdaten an Universitäten.⁹

Zusätzliche universitäre Informationsangebote für den Bereich Datenschutz konnten bis jetzt lediglich auf der Homepage der Universität Ulm gefunden werden. Hier wird mit regelmäßig stattfindenden Weiterbildungen zum Thema Datenschutz geworben.

Zusätzliche Informationsangebote für den Bereich Urheberrecht, mit Links zu weiterführenden Informations- und Schulungsangeboten, unterscheiden sich zwischen den Universitäten hinsichtlich ihres Fokus, ihrer Auffindbarkeit und Aktualität erheblich.

Festzuhalten ist, dass an den Landesuniversitäten sowohl zum Urheberrecht als auch zum Thema Datenschutz Informations- und Beratungsangebote existieren, es sich aber gezeigt hat, dass die Qualität und Quantität der Informationsangebote von Universität zu Universität stark schwankt.

2.1.3.2 Rechtliche Bestimmungen als Forschungshindernisse

Die in 2.1.2.2. [Rechtliche Bestimmungen als Forschungshindernisse](#) erwähnte Forderung nach einer Revision der Schrankenregelung im UrhG lehnt sich an die Auffassung an, die von Experten wie Prof. Spindler und Prof. Dr. Metzger in ihrer Stellungnahme vor dem Ausschuss „Digitale Agenda“ des Deutschen Bundestages vorgetragen / vertreten wurde.

Für das Urheberrecht ist die EU-Richtlinie 2001/29/EG¹⁰ der Rahmen, in dem sich das nationale Recht bewegen kann. Nach diesem Stand dürfen Ausnahmen bei Vervielfältigung

⁷ Auf der Homepage von ZENDAS (<https://www.zendas.de>) ist zu lesen, dass an den Universitätsstandorten Freiburg, Heidelberg, Hohenheim, Konstanz, Mannheim, Tübingen und Ulm es mindestens zwei Ansprechpartner gibt. Eine Person davon mit eher rechtlichem und eine mit eher technischem Hintergrund. Auf den Homepages wurden lediglich die Ansprechpartner mit rechtlichem Hintergrund (Mitarbeiter der ZUV) gefunden.

⁸ Homepage der Seite: <https://www.zendas.de> [letzter Zugriff 30.6.2015].

⁹ Am 30.6.2015 ist bspw. unter den Top-20-Themen der vergangenen Woche der ZENDAS erst auf Platz 17 das erste direkt mit FDM zusammenhängende Thema (Personenbezug bei wissenschaftlichen Umfragen).

¹⁰ <http://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32001L0029&from=DE> [letzter Zugriff 18.6.2015].

und Zugänglichmachung für wissenschaftliche Forschung gemacht werden (Artikel 5(3)a)). Das deutsche Urheberrecht schöpft diesen Rahmen derzeit nicht voll aus. § 53¹¹ und § 57¹² des UrhGs würden sich im Prinzip zwar sehr weit auslegen lassen, jedoch gibt es keine klare Schrankenbestimmung zum wissenschaftlichen Datenaustausch und so bewegen sich Forschende schnell in einem rechtlichen schlecht geschützten Bereich.

Die Datenschutzrichtlinie der EU befindet sich im Augenblick im Wandel. Wir können hier nur den aktuellen (inoffiziellen) Stand¹³ der Vorschläge für eine neue Datenschutzgrundverordnung in Bezug auf Forschungsfragen zusammenfassen, haben aber auch noch einige Inkonsistenzen ausmachen können, sodass nicht klar ist, was genau gelten wird.

1. Kurzfassung Artikel 5e: Löschfristen sollen für Forschung nicht mehr automatisch gelten. Forschende dürfen unter Einhaltung angemessener Schutzmaßnahmen für notwendige Zwecke Daten aufheben dürfen, wenn die Zweckgegebenheit regelmäßig überprüft wird.
2. Kurzfassung Artikel 81 2a: Wissenschaftliche Forschung kann von den Mitgliedstaaten vom Einwilligungszwang bei der Verwendung von medizinischen, personenbezogenen Daten befreit werden. Allerdings ist hier ein großes öffentliches Interesse erforderlich, das noch genauer spezifiziert werden muss.
3. Kurzfassung Artikel 83: Forschung darf mit personenbezogenen Daten stattfinden, wenn a) es nicht anders geht und die Daten so weit es möglich ist anonymisiert sind und b) zusätzliche Zuordnungen von Personen zu den Daten von den Daten (weit) getrennt aufbewahrt sind und auch sonst alle Maßnahmen ergriffen sind, um unbefugte Rückschlüsse auf die Person zu verhindern.

Hier wird sich zeigen, wie in nationalem Recht „alle Maßnahmen“ interpretiert werden und ob daraus insbesondere automatisch immer ein Offline-Zwang des verarbeitenden Systems resultiert.

2.1.3.3 Schutz- und Vermarktungsrechte für produzierte Daten

Schutz gegenüber Verwertern Was die Stärkung der Machtposition der Forschenden gegenüber den Verlagen betrifft, so wurde mit dem Anfang Dezember 2014 in Kraft getretenen Zweitveröffentlichungsrecht¹⁴ von Seiten der Politik ein erster Schritt in diese Richtung unternommen. Das Zweitveröffentlichungsrecht aus § 38, Abs. 4 räumt dem „Urheber eines wissenschaftlichen Beitrags“ das Recht ein, diesen „nach Ablauf von zwölf Monaten seit der Erstveröffentlichung in der akzeptierten Manuskriptversion öffentlich zugänglich zu machen.“ Dies gilt unabhängig davon, ob dem Verleger oder Herausgeber ein ausschließliches Nutzungsrecht eingeräumt wurde. Das zentrale Anliegen dieses Gesetzes (und zwar „die

¹¹ http://www.gesetze-im-internet.de/urhg/__53.html [letzter Zugriff 18.6.2015].

¹² http://www.gesetze-im-internet.de/urhg/__57.html [letzter Zugriff 18.6.2015].

¹³ <http://www.delegedata.de/wp-content/uploads/2014/06/DS-GVO-konsolidiert.pdf> [letzter Zugriff 30.6.2015].

¹⁴ Siehe § 38, Abs. 4 UrhG.

rechtliche Absicherung des »Grünen Weges« zum Open Access) kann aufgrund der mit der Regelung verbundenen Bedingungen jedoch nicht immer realisiert werden.“¹⁵

Schutz gegenüber Kooperationspartnern Bei der Durchsetzung ihrer Interessen gegenüber potentiellen Kooperationspartnern aus der Wirtschaft und Industrie werden die Forschenden bereits jetzt von den jeweils zuständigen Abteilungen für Technologietransfer (ZUV) unterstützt. In welchem Umfang diese Angebote genutzt werden und ob die Angebote den Bedürfnissen der Forschenden gerecht werden, kann an dieser Stelle jedoch nicht beantwortet werden.

Für die Befürchtung mancher Forschender Open Access könnte die Vertrauensbasis für personenbezogene Forschung zerstören, lassen sich momentan keine Anhaltspunkte finden. Gerade bei sensiblen Daten haben auch die Forschenden stets das notwendige Schutzbewusstsein gezeigt und die Standards des Datenschutzes werden auch von Seiten der Politik nicht zugunsten von Open Access aufgeweicht, sodass ein vertrauensvolles Arbeiten wohl auch in Zukunft möglich bleiben wird.

2.1.3.4 IT- Infrastruktur

Austausch von Daten Seit dem 1.1.2015 steht den Forschenden der baden-württembergischen Hochschulen mit bwSync&Share ein (Cloud-) Dienst zur Synchronisation und zum Austausch von Daten zur Verfügung. „Anders als die Dropbox“ soll dieser Dienst „die Anforderungen an Datensicherheit und Datenschutz“ adäquat erfüllen.¹⁶ Zur Speicherung personenbezogener Daten ist dieser Dienst bisher jedoch nicht geeignet.¹⁷

Speicherung und Archivierung Die langfristige physikalische Bewahrung und Pflege der „Bitstreams“ ist eine Aufgabe des Landesspeichers bwLSDF.¹⁸ „Die Large-Scale Data Facility (LSDF) stellt Hardwareressourcen zur Speicherung, Archivierung und Analyse großskaliger Daten aus prinzipiell allen Wissenschaftsbereichen zur Verfügung. Gleichzeitig dient die LSDF als zentraler Datenspeicher für Baden-Württemberg“¹⁹ „Mit dem Begleitprojekt bwLSDF fördert das MWK die Erweiterung der LSDF als zentralen Datenspeicher für die baden-württembergischen Hochschulen.“²⁰ Dieser Dienst ist jedoch ebenfalls bisher nicht für personenbezogene Daten geeignet.

¹⁵ Das Zweitveröffentlichungsrecht des URHG § 38 Abs. 4 – Möglichkeiten und Grenzen bei der Anwendung in der Praxis, <http://epic.awi.de/35491/1/ZUM.pdf> [letzter Zugriff 18.6.2015].

¹⁶ ESFK, 87.

¹⁷ Nutzerhandbuch von bwSync&Share, 44 <http://www.scc.kit.edu/downloads/sdm/Nutzerhandbuch-bwSyncAndShare.pdf> [letzter Zugriff 30.6.2015].

¹⁸ Das Zweitveröffentlichungsrecht des URHG § 38 Abs. 4 UrhG – Möglichkeiten und Grenzen bei der Anwendung in der Praxis, 84, <http://epic.awi.de/35491/1/ZUM.pdf> [letzter Zugriff 18.6.2015].

¹⁹ ESFK, 89.

²⁰ ESFK, 87.

2.1.4 Gegenwärtige Handlungsempfehlungen

2.1.4.1 Klärung von Rechtsfragen und Rechtssicherheit

Es gibt zwar an allen Landesuniversitäten Ansprechpartner für Datenschutz und Urheberrecht, dennoch sind einige Kernbedarfe der FDM-Rechtsberatung oft noch unzureichend gedeckt. Da hierfür vermutlich nicht ein einzelner Grund verantwortlich ist, erscheinen mehrere parallele Maßnahmen mit unterschiedlichem Fokus angebracht.

Informationsmaterial Es wäre wünschenswert existierende Erfolgsmodelle der Umsetzung von Beratungs- und Informationsangeboten zu identifizieren und einer breiten (Fach-)Öffentlichkeit zugänglich zu machen. Das Informationsmaterial der ZENDAS könnte man eventuell aus einer FDM-Web-Plattform heraus so verlinken, dass die Top-Themen des FDMs im Vordergrund stehen.

Führerschein für Datenverarbeitung Die Forschenden sollten nicht nur über Datenschutz, sondern auch über das Recht auf Forschung genügend informiert werden. Die Forschenden brauchen manchmal auch Rückendeckung und Kenntnisse, die ihnen ermöglichen nicht, jede Datensicherheitsempfehlung wortgetreu umsetzen zu müssen, wenn der zusätzliche Schutz gegen ein ohnehin immer existentes Restrisiko die Forschungsabläufe in außerordentlichem Maß behindert, ohne sich für jede zusätzliche Datenverarbeitungsaktion das Einverständnis der ZENDAS besorgen zu müssen. Hier könnte man analog zum Verkehrswesen einen „Führerschein für Datenschutz“ einführen, der zertifiziert, dass ein Forschender das notwendige Wissen zum Datenschutz erworben hat und der einen größeren Handlungsspielraum erlaubt. Das beinhaltet beispielsweise den erlaubten Austausch von personenbezogenen Daten unter zwei zertifizierten Forschenden. Die Maßnahme wäre im Datenschutzgesetz zu verankern.

Zertifikate für Datenschutz der IT-Infrastruktur Analog zum Datenführerschein für Personen, wäre es für viele Forschende hilfreich, Software oder Infrastruktur von IT-Experten für Datenschutz sichtbar zertifizieren zu lassen und solch zertifizierte Software bevorzugt für eine Verarbeitung personenbezogener Daten einzusetzen. Die ZENDAS bietet eine solche Zertifizierung für Hochschulsysteme an. Es könnte mit anderen Bundesländern kooperiert werden, die ähnliche Ansätze verfolgen. Das wäre eine Möglichkeit über Insellösungen, die jeder Datenverarbeitende selbst betreibt (und die vermutlich nicht wesentlich sicherer sind), hinwegzukommen. In jedem Fall ist darüber nachzudenken, wie die bestehenden Landesdatendienste entsprechend erweiterbar sind, um mit ihnen personenbezogene Daten zu verarbeiten.

2.1.4.2 Rechtliche Bestimmungen als Forschungshindernisse

Datenschutz Für den Bereich des Datenschutzes gilt es zu prüfen, inwieweit zum Zwecke der Datenakquise, Big Data oder der Aufbewahrung und Weitergabe von Forschungsdaten im Landesdatenschutzgesetz (LDSG) Ausnahme- bzw. Sonderregelungen für die Forschung geschaffen werden können.

Digitale Buchausleihe Hier sollten die universitären Bibliotheken im Verbund mit anderen großen Bibliotheken durchaus auch E-Science vorantreiben. Einige Verlage verteidigen hier noch immer Pfründe aus einer Buchdruckzeit. So müsste beispielsweise Streaming von urheberrechtlich geschützten Büchern ganz analog zur klassischen Ausleihe mit entsprechenden Vorkehrungen gegen Replikation erlaubt sein (s. dazu auch [4.2. Erwartungen an die Digitalisierung](#)).

Urheberrecht Das Urheberrecht behindert Forschende, die sich nicht trauen, selbst erhobene Daten für andere zur Verfügung zu stellen, wenn andere daran Ansprüche anmelden könnten. Hier sollte §53 UrhG so weit ausgelegt werden dürfen (oder verändert werden), dass das Urheberrecht zeitgemäß bleibt. Das Verwendungsrecht aus §53 UrhG „zum eigenen wissenschaftlichen Gebrauch“ ergibt nur wirklich Sinn, wenn man die Daten auch mit anderen Forschenden besprechen kann und diese sich ebenfalls eine Kopie von den Daten machen dürfen. Das sollte wiederum über Repositorien möglich sein, um keine unnötigen Hürden für die Forschung zu erzeugen.

2.1.4.3 Schutz- und Vermarktungsrechte für Datenproduzenten

Schutz gegenüber Verwertern Es empfiehlt sich, die Forschenden aktiv über ihr Recht zur Zweitveröffentlichung und die hierfür notwendigen Voraussetzungen zu informieren. Darüber hinaus müssen die Forschenden von ihren Universitäten aktiv bei der Durchsetzung dieses Rechts gegenüber Verlagen begleitet und unterstützt werden.

Schutz gegenüber Kooperationspartnern Wie in [2.1.3.3. Schutz- und Vermarktungsrechte für produzierte Daten](#) bereits erwähnt, existieren in den zentralen Universitätsverwaltungen Abteilungen, welche die Forschenden bei der Durchsetzung ihrer Interessen gegenüber potentiellen Kooperationspartnern aus der Wirtschaft unterstützen. Ähnlich wie bei den Beratungsangeboten wäre es interessant herauszufinden, in welchem Umfang diese Angebote genutzt werden und ob die existierenden Angebote in ihrer aktuellen Form den Bedürfnissen der Forschenden gerecht werden.

2.2 Informationsangebote (469)

2.2.1 Einleitung

In diesem Kapitel wird der Frage nachgegangen, wie gut die Forschenden an den baden-württembergischen Institutionen bereits über Forschungsdatenmanagement informiert sind und wie durch interessante Inhalte, auf Forschende zugeschnittene Informationskanäle und übersichtliche Strukturierung von Informationen das Thema einer großen Mehrheit näher gebracht werden kann. Dass Informationen zum FDM für alle Fachrichtungen und Wissenschaftsbereiche gleichermaßen interessant sind, lässt sich aus den Fachrichtungsauszahlungen unserer gesammelten User-Stories ablesen. Da das Thema in der genommenen Stichprobe entsprechend über die Wissenschaftsbereiche verteilt ist (vgl. [Abb. 4](#) mit [Abb. 1](#)).

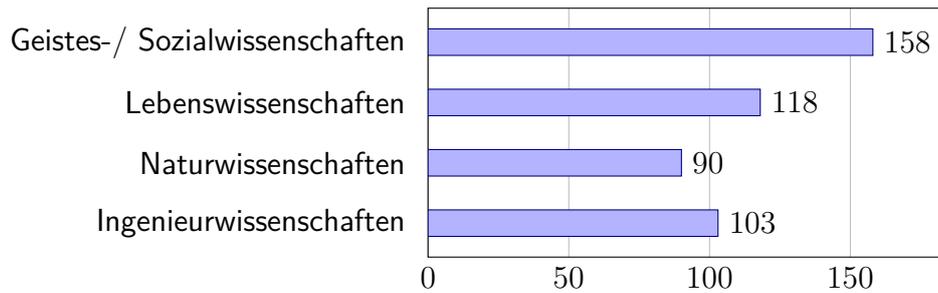


Abb. 4: Verteilung der User-Storys nach der Fachsystematik der DFG

Die Analyse der 469 zum Thema Informationsangebote zugehörigen User-Storys zeigt, dass sich die Universitäten von Baden-Württemberg noch am Beginn der E-Science-Ära befinden. Der Kenntnisstand in den einzelnen Fachgebieten ist sehr unterschiedlich, was die Dringlichkeit einer effizienteren Informationspolitik verdeutlicht.

Zur weiteren Auswertung wurden die Antworten auf die folgenden Fragen des Interviewleitfadens herangezogen:

- Frage 25a: Wie hoch ist der Bedarf Ihrer Forschungsgruppe mehr über Methoden des Forschungsdatenmanagements zu erfahren?
- Frage 25b: Welche Methode der Information würden Sie bevorzugen?
- Frage 26: Was denken Sie, wie gut Sie bisher über (Landes-) Datendienste informiert werden? Wie möchten Sie in Zukunft über existierende oder geplante Services informiert werden?

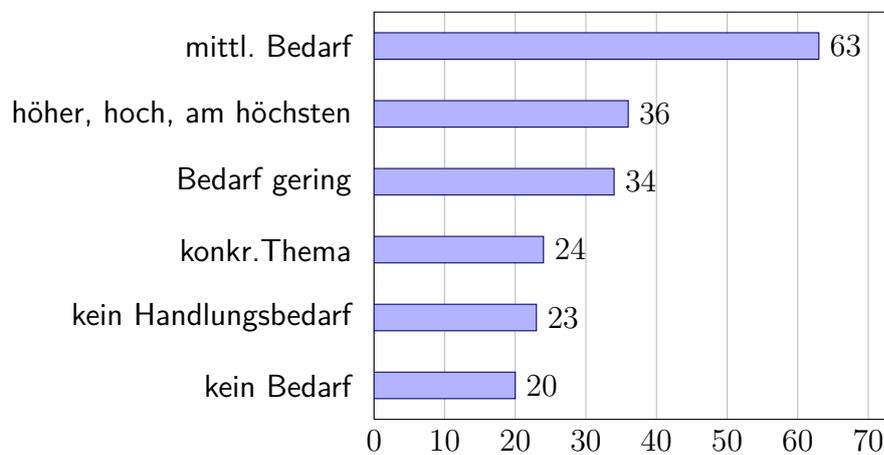


Abb. 5: Antworten auf Frage 25a: „Wie hoch ist der Bedarf Ihrer Forschungsgruppe mehr über Methoden des Forschungsdatenmanagements zu erfahren?“

Allgemein betrachtet, lässt sich aus den Daten in [Abb. 5](#) ablesen, dass von 170 Befragten, die eine Einschätzung abgaben, 99 einen mittleren bis sehr hohen Bedarf an mehr Informationen zum Thema Forschungsdatenmanagement in ihren eigenen Arbeitsgruppen sehen.

Lediglich 54 Befragte gaben an für mehr Informationen nur einen geringen oder gar keinen Bedarf zu haben. 24 Befragte haben auf die Frage ein konkretes Thema des FDMs genannt für das sie sich Informationen wünschen, ohne dass damit zusammen eine Bedarfshöheneinschätzung abgegeben wurde. Etwa gleich viele (23) gaben an zwar einen Bedarf zu sehen, wollen sich aber aktuell nicht tiefer mit FDM zu beschäftigen.

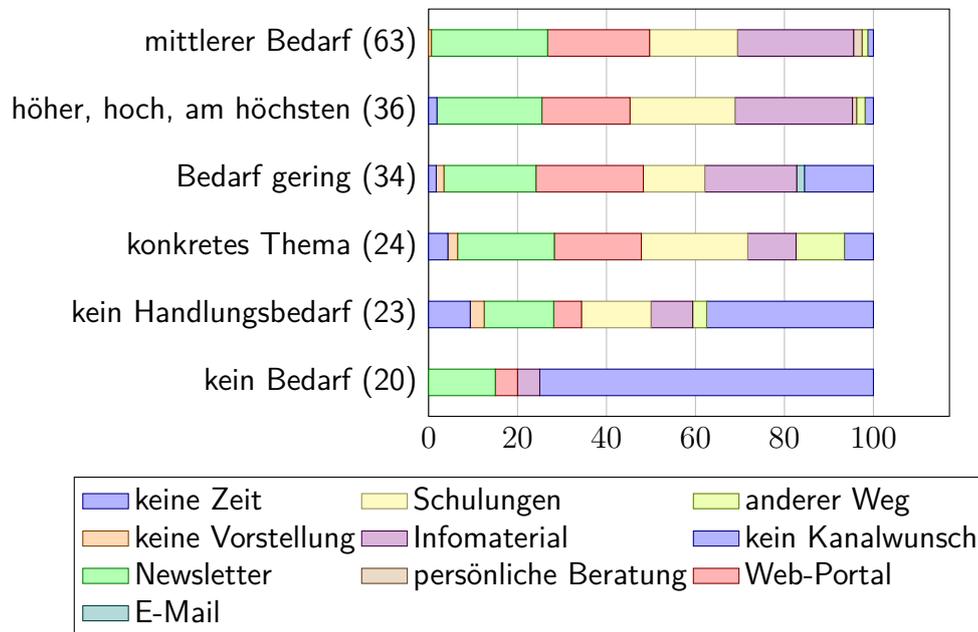


Abb. 6: Relative Verteilung der Informationskanalnennungen (25b) bei vorhandener Antwort auf 25a. Die Zahl in Klammern gibt die Gesamtmenge der Nennungen an.

Abb. 6 zeigt die gewünschte Art der Informationskanäle für verschiedene Dringlichkeitsstufen der Information. Man erkennt, dass vorhandene Bedarfe auch in vermehrten Kanalwünschen münden und mit höherer Dringlichkeit dann auch Zeit zur Verfügung steht sich zu informieren. Nicht auf der Grafik zu sehen ist, dass mit höherer Dringlichkeit auch immer mehr Kanäle genannt wurden. Auffällig ist, dass vor allem bei denen die keine Bedarfshöhe festgelegt haben, aber Informationen zu einem konkreten Thema brauchen, Schulungen und „andere Wege“ gehäuft sind. Schulungen scheinen vor allem bei intensiver Beschäftigung (hoher Bedarf) öfters genannt zu werden. Die „anderen Wege“ beschreiben mehrheitlich Talks/Präsentationen zu einem Thema oder Online-Lehrveranstaltungen.

Noch deutlicher ist der Informationsbedarf beim Thema (Landes-) Datendienste (s. Abb. 7). Hier fühlen sich etwa zwei Drittel der Forschenden gering oder gar nicht über die Möglichkeiten informiert. Diese beiden Ergebnisse vermögen angesichts der Tatsache, dass an den Hochschulen originäre, FDM-spezifische Informationsangebote bisher nur vereinzelt zu finden sind, kaum zu überraschen.

Erste Anstrengungen mehr Informationen zum Thema bereitzustellen, sind bereits unternommen worden. Von diesen Bemühungen konnten offensichtlich einige der Forschenden profitieren, die sich bereits gut bis sehr gut informiert fühlen. Dabei handelt es sich um etwa jeden vierten Forschenden. Auffällig ist jedoch, dass ein Großteil dieser Gruppe schon

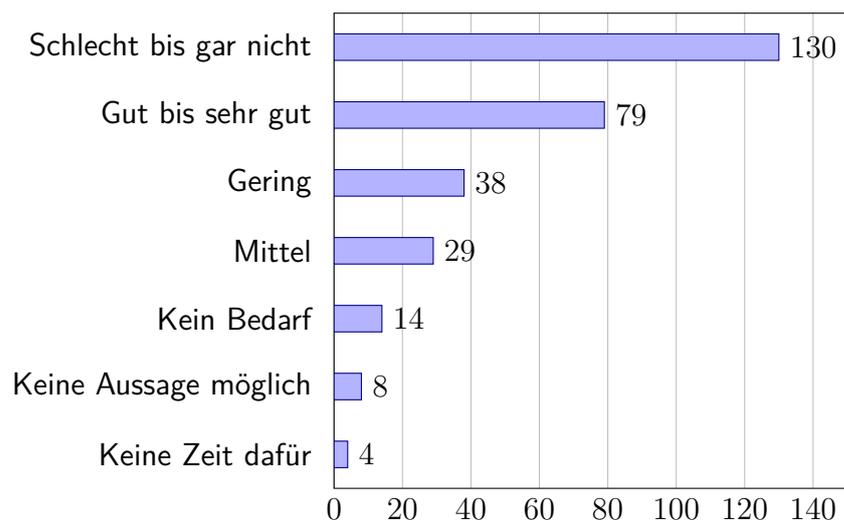


Abb. 7: „Was denken Sie, wie gut Sie bisher über (Landes-)Datendienste informiert werden?“ (Frage 26)

vor dem Interview aus verschiedenen Gründen ein hohes Interesse an FDM hatte. Es gilt also vor allem die Gruppen zu erreichen, zu denen bisher wenige bis gar keine Informationen vorgedrungen sind bzw. die Relevanz der Informationen zu verbessern und an den spezifischen Bedarf dieser Zielgruppe anzupassen. Die Anzahl derer, die bei sich und der eigenen Arbeitsgemeinschaft keinen Bedarf an Informationen sehen, ist mit 5 % relativ gering. Auch dies verdeutlicht, dass das Informationsdefizit nicht einem generellen Desinteresse am Thema Forschungsdatenmanagement geschuldet ist.

2.2.2 Bedarf der Wissenschaftsgemeinschaft

Was den Bereich Informationsangebote angeht, konnten wir zwischen den Empfehlungen des ESFKs und den Wünschen der Forschenden einen hohen Grad an Übereinstimmung feststellen (s. 2.2.4. *E-Science-Bestrebungen*). Die konkret geäußerten Wünsche der Forschenden sind naturgemäß vielfältiger und detaillierter als die Aussagen des ESFKs, bündeln sich aber gut in den dort angeschnittenen Themen. In den folgenden Abschnitten werden die Wünsche der Forschenden verdichtet wiedergeben. Ebenso wird kurz aufgezeigt, wo der Status quo diese Wünsche bereits abbildet und an welchen Stellen noch größere Diskrepanzen bestehen. Wie bereits in der Einleitung angedeutet, ist das Interesse an Informationen zum Thema Forschungsdatenmanagement unter den Forschenden relativ hoch.

Aus den Interviews lassen sich zwei Trends abzulesen. Erstens wünscht sich eine Mehrheit der befragten Personen mehr Informationen zum Forschungsdatenmanagement. Zum Zweiten gibt es nicht den einen klaren Favoriten bei den Informationskanälen, sondern eine Vielzahl an Wünschen, die bereits vermuten lassen, dass ein multimodales Vorgehen den größten Erfolg in der Verbesserung der Informationslage zeigen würde. Im Folgenden werden die von den Forschenden favorisierten Informationswege und Angebote noch einmal näher beleuchtet. Basis der Analyse bilden hierbei die User-Stories.

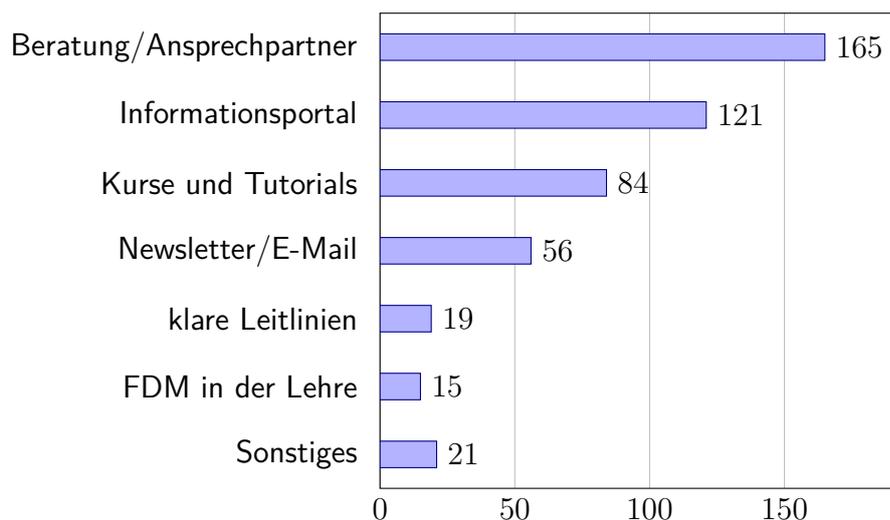


Abb. 8: Bedarfe im Bereich Informationsangebote

2.2.2.1 Beratung / Ansprechpartner (165)

„Als Forschende, wünschen wir uns persönliche Beratung für unsere FDM-Fragen vor Ort oder auch telefonisch, um schnell und persönlich Hilfe und Erklärung zu unseren Fragestellungen zu erhalten.“

Mit 165 User-Stories rangiert der Wunsch nach einer Beratungsmöglichkeit in irgendeiner Form (per Online-Support, E-Mail, Telefon oder persönlich) an erster Stelle. Einen solchen Service stellen sich die meisten Befragten als permanente Einrichtung vor. Nur vier Befragte gaben an, sie würden eine Beratung nur bei Bedarf oder nur in zeitlich festgelegten Intervallen nutzen wollen. In 28 dieser User-Stories wünschen sich die Forschenden explizit eine persönliche Beratungsstelle vor Ort, also an der eigenen Universität. Ausschlaggebend für diesen Wunsch war, dass in einem Vier-Augen-Gespräch Sachverhalte und Probleme nach dem Eindruck der Befragten schneller verständlich gemacht werden können als am Telefon oder per E-Mail. In einigen User-Stories wird explizit der Wunsch nach der Einrichtung eines eigenen FDM-Beauftragten zum Ausdruck gebracht.

Inhaltlich gesehen ist der Wunsch nach einer direkten FDM-Beratung am häufigsten ausgeprägt, gefolgt vom Wunsch nach technischer Beratung im FDM-Bereich. In einigen User-Stories wünschen sich die Befragten zudem spezifisch Beratung zu vorhandenen (Landes-)Datendiensten.

2.2.2.2 Informationsportale (121)

„Als Forschende, wünschen wir uns eine Webplattform, die allgemeine Informationen zum Forschungsdatenmanagement, Informationen zu bestehenden und neuen Datendiensten sowie Tutorials und Anleitungen zu FDM-Themen enthält, um uns selbst bei Bedarf informieren zu können.“

Eine besonders große Diskrepanz zwischen Angebot und Nachfrage besteht beim Thema Informationsportale. In 121 User-Stories wird der Wunsch nach einer solchen Plattform im

Internet geäußert. Allein ein Drittel der User-Stories hier haben den Wunsch nach allgemeinen Informationen zu Definitionen und Methoden des Forschungsdatenmanagements zum Inhalt. Ein weiteres Viertel der Befragten wünschen sich ein Informationsportal, auf welchem alle bestehenden, neuen und in Planung befindlichen Datendienste der Universitäten und des Landes übersichtlich zusammengefasst sind. Etwa jeder Zehnte wünscht sich eine Plattform mit einer Liste aller Forschungsvorhaben, also ein sogenanntes Forschungsinformationssystem. Hierbei wünschen sich in etwa die Hälfte dieser User-Stories die Interviewten ein lokales, universitäres System und die andere Hälfte ein zentrales, landesweites System. Auch gab es vereinzelte Wünsche nach sehr speziellen Informationsplattformen wie beispielsweise einem Portal für Softwareempfehlungen, Datenbankerstellung oder Gerätwartung. Generelle Anforderungen an eine Webplattform sind Aktualität und Übersichtlichkeit. Weiterhin sollte ein solches Portal leicht zu durchsuchen sein.

2.2.2.3 Kurse und Tutorials (84)

„Als Forschende, wünschen wir uns für uns maßgeschneiderte Schulungen und Tutorien zu allgemeinen Themen des Forschungsdatenmanagements, um einen Grundstein für gutes Forschungsdatenmanagement zu legen.“

Auch der Wunsch nach Veranstaltungen wie Workshops, Schulungen oder Seminaren ist groß. Insgesamt gibt es 84 User-Stories, die den Wunsch nach Kursen im Bereich FDM thematisieren. Davon wünschen sich 26 Befragte ein größeres Angebot an allgemeinen FDM-Schulungen. 13 weitere Forschungsgruppen würden FDM-Einführungskurse speziell für Neuestellte (sowohl für wissenschaftliches als auch für IT-Personal) sowie speziell für junge Forschende begrüßen. Außerdem kommt in 15 User-Stories der Wunsch nach einem besseren Schulungsangebot für im FDM-Bereich genutzter Software zum Ausdruck. Drei interviewte Personen sprachen von sich aus an, dass es für Zielgruppen mit besonderen Bedarfen wie in den Geisteswissenschaften oder mit besonderer Verantwortung wie dem Leitungspersonal speziell zugeschnittene Kurse geben sollte. Häufige Anforderungen an ein solches Weiterbildungsangebot sind Praxisorientiertheit und Themenvielfalt sowie der Wunsch einen Kurs besuchen zu können, der dem eigenen Kenntnisstand entspricht.

2.2.2.4 Newsletter / E-Mails (56)

„Als Forschende, wünschen wir uns einen Newsletter, um uns zu Themen des Forschungsdatenmanagements übersichtlich auf dem Laufenden zu halten und über weiterführende Links Details erhalten zu können.“

Ein weiterer gewünschter Informationskanal in den User-Stories zum Thema Informationsangebote war mit 56 Nennungen der Wunsch nach einer Information per E-Mail, konkreter nach einem Newsletter. Aus User-Stories und dem Feedback der Forschenden konnten wir entnehmen, dass meistens ein regelmäßiger FDM-Newsletter nach dem Vorbild des (einseitigen) bwFDM-Projektnewsletters gemeint ist. Als wichtige Anforderungen an einen solchen Newsletter wurden genannt, dass dieser Informationen über bestehende und geplante Datendienste, Services und Tools enthält, nicht häufiger als vierteljährlich erscheint, kurz und prägnant ist sowie bei Bedarf per Link auf weiterführende Informationen verweist. Einen

Sonderfall bildet der Standort Konstanz, wo sich viele Forschende wünschen, solche Informationen in den bereits bestehenden und gut etablierten Newsletter „Forschungsnachrichten“ zu integrieren, um so die Anzahl an zu lesenden Newsletter zu verringern.

2.2.2.5 Leitlinien und Policys (19)

„Als Forschende, wünschen wir uns einheitliche Richtlinien, um klare Anleitung zu bekommen, wie wir mit Forschungsdaten umzugehen haben, die uns dabei aber auch Ermessensspielräume lassen.“

Zum Thema Leitlinien und Policys geht aus den Daten hervor, dass in 19 User-Stories der Wunsch nach Leitlinien oder Vorgaben im Bereich Forschungsdatenmanagement zum Ausdruck gebracht wird. Dabei reicht die Spanne von Richtlinien zur Datenarchivierung und -dokumentation über Richtlinien zu Datenveröffentlichungen bis hin zu einheitlichen Datenschutzrichtlinien. Ebenfalls wünschten sich einige befragte Personen das Aufsetzen einer FDM-Policy ähnlich wie sie bereits an der Universität Heidelberg existiert. Hierbei war den Gesprächspartnern jedoch vor allem wichtig, dass die Richtlinien klare Handlungsvorgaben enthalten, also ein hohes Maß an Detailliertheit erreicht wird. Ebenso sollten solche Richtlinien für alle gleichermaßen verbindlich, also einheitlich, sein. Des Weiteren wurde oft betont, dass eine solche Policy sich an bereits existierenden Standards und Richtlinien der Drittmittelgeber orientieren sollte, um Anforderungskollisionen zu vermeiden.

2.2.2.6 Forschungsdatenmanagement in der Lehre (15)

„Als Forschende, wünschen wir uns, dass kommende Forschungsgenerationen bereits während ihres Studiums mit dem Thema FDM vertraut gemacht und in grundlegenden Techniken und Fertigkeiten geschult werden, um ihre Sensibilität zu erhöhen sie auf Aufgaben in diesem Bereich vorzubereiten.“

In 15 User-Stories wird das Problem angesprochen, dass junge Forschende während der Promotion oder der Postdoc-Forschung häufig mit Datenmanagement-Anforderungen konfrontiert werden, auf die sie in keiner Weise durch ihr vorangegangenes Studium vorbereitet sind. Diese Problematik wird auf einen Mangel an Lernangeboten während des Studiums zurückgeführt. Demzufolge wünschen sich neun der interviewten Communitys, dass Forschungsdatenmanagement als Teil der Grundausbildung eines jeden Studierenden in die Curricula eingeführt werden und weitere drei, dass eine solche Ausbildung Bestandteil eines Doktoratsstudiums sein sollte. Hierbei geht es vor allem um eine Vermittlung grundlegender Kenntnisse, die verbindlich und solide sein soll. Ultimativ erhoffen sich einige der befragten Personen, dass dadurch ein größeres Bewusstsein für die Wichtigkeit der Thematik entsteht und letztendlich ein Einstellungs- und Kulturwandel bewirkt sowie grundlegendes „Handwerkszeug“ für den Umgang und die Arbeit mit Forschungsdaten vermittelt wird. (s. [2.3. Wissenschaftliche Datenkultur](#)).

2.2.2.7 Sonstiges (21)

„Als Forschende, wünschen wir uns einen Informationskanal, welcher uns regelmäßig über bestehende und neue Datendienste und weitere FDM-Neuerungen aufklärt und informiert.“

In 21 weiteren User-Stories nennen die Gesprächspartner weitere mögliche Informationskanäle wie einen Newsfeed (9), Infoblätter / Flyer (2), Vorträge (1) oder ein FDM-Blog (1), während einige schlicht an „mehr Informationen“ im Allgemeinen interessiert sind, ohne einen präferierten Informationskanal nennen zu können.

2.2.3 Status quo

Wie bereits in der Kapiteleinführung angedeutet, ist der Stand der Informations- und Beratungsangebote zum Thema FDM im Land Baden-Württemberg noch ausbaubar. Zwar findet man an allen Universitäten entweder in den Universitätsbibliotheken, Medienzentren, Forschungsdezernaten, Rechenzentren oder Justizariaten Informationen oder Beratungsstellen zu Themen wie Tools, Publikations- und Arbeitsrecht oder Projektbeantragung und -abwicklung. Informationsangebote direkt zum Thema FDM, bilden jedoch bisher die absolute Ausnahme. Im Folgenden werden die Unterpunkte *Leitlinien und Policys, Informationsportale, Kurse und Tutorials* sowie *persönliche Beratung* in Hinblick auf den existierenden Status quo näher betrachtet.

2.2.3.1 Leitlinien und Policys

Von den neun am Projekt teilnehmenden Universitäten hat bisher nur die Universität Heidelberg eine eigenständige Forschungsdatenpolicy herausgebracht.²¹ An der Universität Karlsruhe wird momentan daran gearbeitet, eine Forschungsdatenpolicy mit allen relevanten Gremien abzustimmen. Dadurch sollen die bereits existierenden KIT-Leitlinien „Zur Sicherung guter wissenschaftlicher Praxis“²² ergänzt werden. An der Universität Stuttgart existiert eine „Satzung zur Sicherung der Integrität und zum Umgang mit Fehlverhalten in der Wissenschaft“²³, welche in Abschnitt 5 den Umgang mit Primärdaten in Grundzügen regelt. Die sechs weiteren Hochschulen im Land Baden-Württemberg verfügen nach unserem aktuellen Kenntnisstand nicht über eine solche Leitlinie, noch befindet sich eine solche momentan in Planung. Dies könnte sich jedoch aufgrund der Vorgaben des ESFKs und den dazugehörigen Ausschreibungen in naher Zukunft ändern.

²¹ Research Data Policy, <http://www.uni-heidelberg.de/universitaet/profil/researchdata/> [letzter Zugriff 13.5.2015].

²² Sicherung guter wissenschaftlicher Praxis am KIT, http://www.uni-stuttgart.de/zv/bekanntmachungen/bekanntm_62_2013.pdf [letzter Zugriff 13.5.2015].

²³ Satzung der Universität Stuttgart zur Sicherung der Integrität wissenschaftlicher Praxis und zum Umgang mit Fehlverhalten in der Wissenschaft, http://www.unistuttgart.de/zv/bekanntmachungen/bekanntm_62_2013.pdf [letzter Zugriff 13.5.2015].

2.2.3.2 Informationsportale

Zu vielen Themen existieren im akademischen Umfeld informative Webseiten. Im Rahmen des Themas Forschungsdatenmanagement ist das vorhandene Angebot an Onlineresourcen der Universitäten jedoch noch recht dürftig. Ein Informationsportal, das sich direkt mit dem Thema Forschungsdatenmanagement auseinandersetzt, wird bisher nur auf den Seiten der Universitätsbibliothek des KIT angeboten.²⁴ Im weiteren Umfeld der Thematik ist zudem das Portal [open-access.net](http://www.open-access.net) – der freie Zugang zu wissenschaftlichen Informationen²⁵ zu verorten, welches von der Universität Konstanz in Kooperation mit einigen anderen deutschen Universitäten betrieben wird und zum Thema Open Access mit verschiedensten Informationsmöglichkeiten in Form von Videos, Podcasts, Folien, Flyer und einem FAQ-Segment aufwartet. Die Bibliothek der Universität Freiburg bietet ferner einen kurzen Überblick zum Thema E-Science auf ihren Seiten.²⁶ Des Weiteren existieren an der Universität Tübingen sowie an der Universität Konstanz bereits Forschungsinformationssysteme, welche Informationen zu an diesen Einrichtungen laufenden wissenschaftlichen Projekten bieten. Mit Stand vom 08.01.2015 hat sich auf Landesebene zudem eine Projektgruppe, aus den Universitäten Hohenheim, Stuttgart, Ulm und dem KIT, gebildet, die an diesen Einrichtungen den Aufbau eines Forschungsinformationssystems vorantreiben will.²⁷

2.2.3.3 Kurse und Tutorials

Bei diesem Thema ist die Angebotslage besonders dünn. Nur die Universität Hohenheim stellt Kurse oder Tutorials im Themenumfeld des FDMs in der umfangreichen Video-Tutorial-Sammlung [video2brain](http://www.kim.uni-hohenheim.de/video2brain)²⁸ bereit. Ein kleiner Teil dieser Sammlung ist gratis für Forschende und Studierende anderer Einrichtungen nutzbar. Die Themenbreite reicht von Datenbearbeitungsthemen über Software-Tutorials bis hin zu IT-Fragen. Weitere umfangreiche Onlinekurse oder gar Face-to-Face-Veranstaltungen an den anderen Hochschulen zum Thema FDM sind uns nicht bekannt.

2.2.3.4 Beratung / Ansprechpartner

Neben grundsätzlichen Informationsmöglichkeiten an den Universitätsbibliotheken haben bisher drei Universitäten eine Stelle für persönliche Beratung eingerichtet. Dazu zählen die Universitäten Heidelberg, Mannheim und Tübingen. Am Standort Tübingen richtet sich das E-Science-Center vor allem an die Geisteswissenschaften, die Unterstützung in Fragen der E-Humanities benötigen, während in Mannheim vor allem (aber nicht ausschließlich) Wirtschafts- und Sozialwissenschaften im Fokus stehen. Eine fachübergreifende Einrichtung,

²⁴ Forschungsdatenmanagement, <http://www.bibliothek.kit.edu/cms/forschungsdatenmanagement-journal-policies.php>, [letzter Zugriff 13.5.2015].

²⁵ Open Access: Der freie Zugang zu wissenschaftlicher Information, <http://www.open-access.net/startseite> [letzter Zugriff 13.5.2015].

²⁶ E-Science, <http://www.ub.uni-freiburg.de/index.php?id=escience> [letzter Zugriff 13.5.2015].

²⁷ Projekt Universitätsbenchmark, <http://www.uni-ulm.de/einrichtungen/zuv/qm/akademisches-berichtswesen/projekte.html> [letzter Zugriff 13.5.2015].

²⁸ Onlinekurse mit video2brain, <https://kim.uni-hohenheim.de/98727> [letzter Zugriff 13.5.2015].

die Services rund um das Thema Forschungsdaten anbietet, gibt es in Heidelberg. Während das „Forschungsdaten Service Center“ in Mannheim vor allem zum Thema IT berät, bietet das Zentrum in Heidelberg auch Informationen zu rechtlichen, finanziellen und organisatorischen Fragen an. An der Universität Ulm und dem KIT befinden sich gebündelte Beratungsdienste zum FDM aktuell in Planung.

2.2.3.5 Sonstiges

Weitere denkbare und von den Befragten gewünschte Informationsangebote wie Newsletter, RSS-Newsfeeds oder Blogs zum Thema Forschungsdatenmanagement werden bisher an den Hochschulen Baden-Württembergs nicht angeboten. Weiterhin scheint das Thema auch in der Lehre, ob studienbegleitend oder gar als eigener Studiengang, bisher kaum eine Rolle zu spielen.

2.2.4 E-Science-Bestrebungen

Das ESFK listet einige Empfehlungen auf, die sich auf Informationsstrategien und/oder Informationsangebote beziehen. So werden bspw. die Themen „Leitlinien und Policys“, „Forschungsdatenmanagement in der Lehre“, „Weiterbildung Berufstätiger“, sowie „Beratung“ in konkreten Vorschlägen aufgegriffen.

Zum Thema „Leitlinien und Policys“ sieht das ESFK vor, dass Universitäten Leitlinien entwickeln und vereinbaren, welche die Aufbewahrungsdauer von Forschungsdaten regeln.²⁹ Dies wurde an einzelnen Einrichtungen bereits umgesetzt oder befindet sich in Vorbereitung. Ebenso wird angeregt, landesweite Policys zur Langzeitarchivierung zu entwickeln und umzusetzen.³⁰ Eine landesweit abgestimmte Policy ist bisher nicht angedacht worden. Hierzu bedarf es gemeinsamer Überlegungen aller beteiligten Akteure einschließlich der Hochschulen, des MWK und optimalerweise Vertretern der Förderorganisationen (s. 2.2.5. [Gegenwärtige Handlungsempfehlungen](#)). Sinnvoll wäre es, wenn das MWK für einen solchen Schritt die Federführung übernehme.

Aufgegriffen wird im Fachkonzept ebenso die bisher fehlende Vermittlung des Themas Forschungsdatenmanagement und der entsprechenden Kompetenzen in den Curricula vor allem forschungsdatenintensiver Studiengänge. Hierzu heißt es:

„Das Thema Forschungsdaten kommt i.d.R. gar nicht in der Lehre vor. Im Hinblick auf die Anforderungen in nächster Zukunft muss Datenmanagement Teil der notwendigen Ausbildung im Bereich Forschungsmethoden werden. Der sichere und nachvollziehbare Umgang mit Forschungsdaten über den gesamten Data Life Cycle wird zunehmend in allen Wissenschaftsbereichen als grundlegende Kompetenz benötigt.“ (ESFK, 88)

Es soll eine verbesserte Informationslage junger Forschender und mehr Kompetenz im Umgang mit Forschungsdaten sowohl durch Lehrangebote, die sich an eine breite Zielgruppe

²⁹ ESFK, 89.

³⁰ ESFK, 90.

richten, als auch durch disziplinspezifische Kurse, die für bestimmte Fächer wichtige Schlüsselqualifikationen vermitteln, erreicht werden. Ebenso wird empfohlen, die Kursinhalte der jeweiligen Qualifizierungsstufe der Adressaten (Bachelor, Master, Promotion etc.) anzupassen.³¹ Bisher spielt dieses Thema in den Studiengängen der baden-württembergischen Universitäten überhaupt keine Rolle. Hier besteht die größte Diskrepanz zu den Empfehlungen des Fachkonzepts. Eine Empfehlung bezüglich des Lehrmodus (Online-Tutorials, Selbstlernkurse, klassisch *Face-to-Face*) wird nicht abgegeben.

Neben studienfachbegleitenden Modulen sieht das ESFK zudem die Einführung eines Studiengangs „Data Scientist / Data Librarian“ vor.³² Auch in diesem Bereich gibt es bisher keine erkennbaren Ansätze an den Hochschulen Baden-Württembergs.

Weiterhin wird angeregt für bereits Berufstätige, Kenntnisse „in *berufsbegleitenden E-Learning-Kursen*“ zu vermitteln.³³ Leider gibt es auch im Weiterbildungsbereich bisher kaum Angebote.

Mit Hinblick auf das Thema Beratungsangebote zum Forschungsdatenmanagement wird im ESFK die Einrichtung einer landesweiten Koordinationsstelle angesprochen. Zum einen wird eine Clearingstelle für rechtliche Fragen, Lizenzen und Publikationsformen empfohlen³⁴, zum anderen wird die Einrichtung einer landesweiten Beratungsstelle angeregt, welche Hilfe zu Forschungsdatenstandards und der persistenten Adressierung von Dokumenten bieten sowie die Verantwortung für die benötigten Wissenschaftsplattformen übernehmen soll.³⁵ Da bereits an einigen Hochschulen FDM-Beratungsstellen existieren, ist zu überlegen, wie diese in ein solches Angebot integriert werden oder stattdessen zu einem landesweit nutzbaren Service ausgebaut werden können.

2.2.5 Gegenwärtige Handlungsempfehlungen

Basierend auf den hier vorgestellten Erkenntnissen ergibt sich als übergeordnete Handlungsempfehlung die Erarbeitung einer integrierten, gemeinsamen Informationsstrategie des MWK und der Hochschulen. Aus einer solchen gemeinsamen Vorgehensweise würden sich Synergien ergeben, da die Informationslage zum Umgang mit Forschungsdaten an allen Universitäten ähnlich ist. Im Kontext dieser Leitempfehlung stehen daher alle unten aufgeführten Maßnahmen, die eine solche Informationsstrategie mit Leben füllen können.

2.2.5.1 Informationsportal und Beratungsservice

Ein umfassendes, zentrales Web-Informationsportal für FDM-Themen, das sowohl einen Überblick über vorhandene Dienste und Methoden als auch Updates zu Neuigkeiten und hilfreiche Tools bietet, sollte eingerichtet werden. Hierbei sollte besonderes Augenmerk auf Nutzerfreundlichkeit gelegt werden: Sichergestellt werden könnte dies durch sogenannte Usability-Studien oder einstiegfreundliche Feedbackmöglichkeiten auf dem Web-Portal. Bei der Eta-

³¹ ESFK, 89.

³² ESFK, 89.

³³ ESFK, 88.

³⁴ ESFK, 90.

³⁵ ESFK, 90.

blierung einer solchen Plattform sollte darauf geachtet werden, bereits vorhandene Portale und Plattformen wie <http://open-access.net/> zu verknüpfen, um einen großen Informationseffekt mit kleinem Aufwand zu erzielen. Um dies zu realisieren, wäre die Etablierung eines zentralen FDM-Teams, welches das Portal betreibt und einen alle 3–6 Monate erscheinenden Newsletter herausbringt, sinnvoll. Dasselbe Team könnte ebenso Ansprechpartner für die persönliche Beratung und für Entwicklungsprojekte, die den Bedarf der Forschenden berücksichtigen sollen, sein. Eine solche Beratung kann telefonisch erfolgen und über ein Ticketsystem auf dem Informationsportal organisiert werden. Dies hätte den zusätzlichen Vorteil, dass Use-Cases gesammelt und nach einem gewissen Zeitraum zu einem FAQ-Segment zusammengefasst werden können.

2.2.5.2 Kurse, Workshops, Tutorien

Hilfreich wäre die Anregung eines breiteren Informationsangebots durch die Universitätsbibliotheken, Rechenzentren, Medienzentren oder FDM-Beauftragten der Universität. Dieses Weiterbildungsangebot könnte in Form von Workshops, Webinars, Online-Tutorien, und Inhouse-Präsentationen erfolgen.

2.2.5.3 Richtlinien und Policys

Eine Vorlage für eine FDM-Richtlinie sollte entwickelt und eine Empfehlung der Übernahme und Umsetzung an die Hochschulen (Orientierung an Vorgaben der DFG / EU) ausgesprochen werden. Ideal wäre eine gemeinsame Erarbeitung und Umsetzung einer solchen Richtlinie. Um Anreize dafür zu setzen, empfiehlt es sich, die Existenz einer solchen Leitlinie als Kriterium für die Vergabe von Projekten im E-Science-Programm des MWK festzulegen.

2.2.5.4 Forschungsdatenmanagement in der Lehre

Um das Thema Forschungsdatenmanagement in der Lehre zu etablieren, wäre eine Empfehlung des MWK an die Hochschulen, in datenintensiven Disziplinen Grundzüge des Themas Forschungsdatenmanagement (Datenlebenszyklus, OAIS) in die Curricula aufzunehmen, sinnvoll. Ebenso sollte eine solche Empfehlung den Hinweis beinhalten, Lehrangebote den fachspezifischen Bedarfen der Forschenden (beispielsweise „Anonymisierung“ in der Psychologie und der Humanmedizin oder „Rechnernutzung“ in den Digital Humanities) anzupassen.

2.3 Wissenschaftliche Datenkultur (199)

2.3.1 Einleitung

In den Interviews wurde von vielen Forschenden der Wunsch nach einem Anreizsystem für die Publikation von Forschungsdaten und ein tiefgehender Wandel in der Handhabung von Daten gefordert. Von der Erhebung über die Analyse bis zum Teilen der Daten wird eine offenere Einstellung von Forschenden und die Unterstützung dieser Abläufe durch die zentralen universitären Einrichtungen erwartet. Die Einstellung zur Wichtigkeit von hochwertigen

Daten soll sich diesen User-Stories zufolge grundlegend ändern. Dieser Wandel betrifft alle Personen und Organisationen die mit Daten in der Wissenschaft in Berührung kommen.

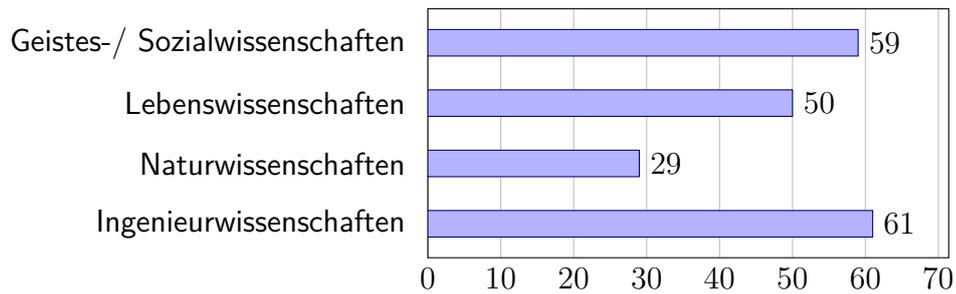


Abb. 9: Verteilung der User-Stories nach der Fachsystematik der DFG

In unseren Interviews haben wir die Frage gestellt, ob es Datenformate im jeweiligen Forschungszweig gibt, mit denen die Forschenden unzufrieden sind (und falls ja, welche diese sind). Die häufigsten Probleme haben die Forschenden mit ASCII, Microsoft (mit Word und Excel), PDF und gerätespezifischen Datenformaten.

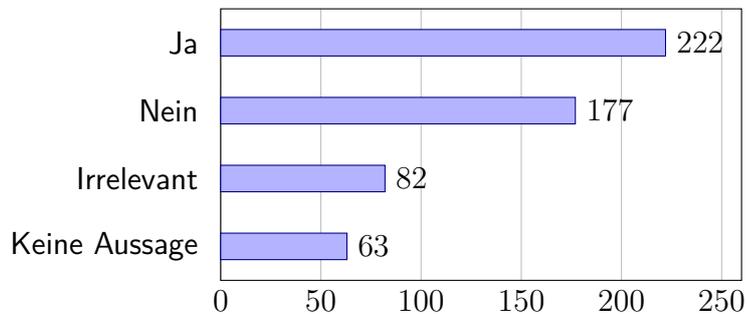


Abb. 10: „Sind Sie zufrieden mit den öffentlich zugänglichen Datenquellen?“ (Frage 21)

Aus den Antworten wird deutlich, dass die Forschenden einiger Disziplinen mit der aktuellen Situation nicht zufrieden sind. In einzelnen Gesprächen sind über die Diskussion zum Thema „Nicht-Veröffentlichung von Negativbefunden“ dabei sogar superlative Zustandsbeschreibungen wie „eine einzige Katastrophe“ oder „wahnsinnige Ressourcenverschwendung“ vermittelt worden. Von den Förderorganisationen und Behörden werden klare und konkret befolgbare Richtlinien zur Datenveröffentlichung, -speicherung, -archivierung und Nachvollziehbarkeit von wissenschaftlichen Arbeiten erwartet. Anreizsysteme sollen alle Forschenden dazu bewegen, eine offenere Wissenschaft durch die Verwendung der Prinzipien des Open Access für Ergebnisse, Open Data für Datenpublikationen und Open Source für Programme zu praktizieren. Die Forschenden sehen ein, dass nicht alle Daten veröffentlicht werden können, sei es aufgrund von Auftragsforschung oder der Sensibilität der Daten, jedoch betrifft dies nur einen Teil der Disziplinen.

Um diese Anforderungen zu erfüllen und ein Umdenken zu bewirken, sollen Anreizsysteme geschaffen werden. Im Folgenden wird auf die Wünsche der Forschenden zur generellen Datenkultur eingegangen. Am Ende des Kapitels werden auch Vorschläge für Anreizsysteme gegeben, um den gewünschten Kulturwandel zu initiieren.

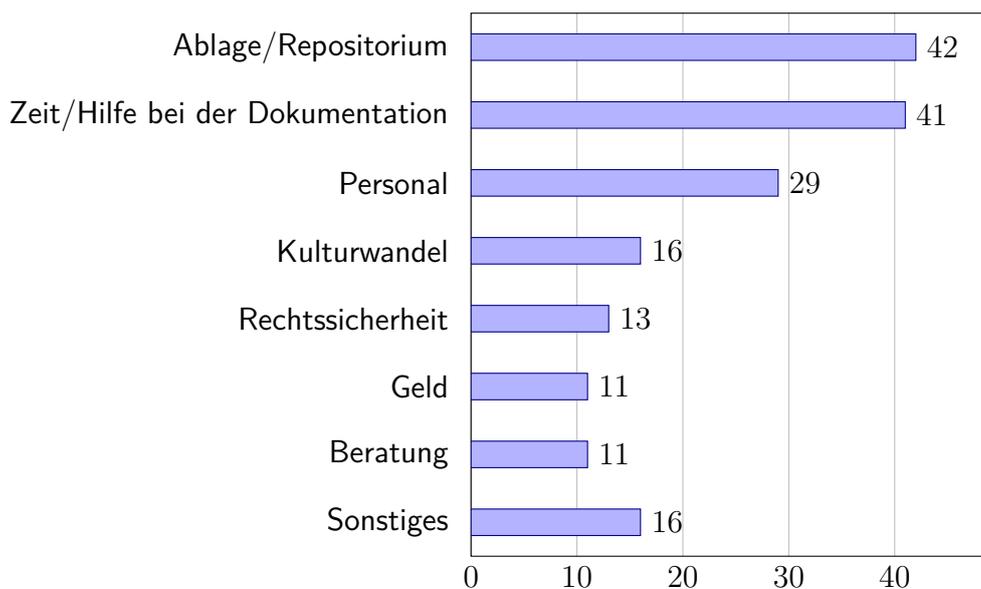


Abb. 11: „Was würden Sie noch brauchen, um Ihre Daten effizient und einfach veröffentlichen zu können?“ (Frage 24)

2.3.2 Bedarf der Wissenschaftsgemeinschaft

Die größten Sorgen der Forschenden bei der Veröffentlichung von rechtlich unproblematischen Datensätzen sind die Zeit und das persönliche Risiko. Die qualitativ hochwertige Aufbereitung der Daten, sodass sie von anderen verstanden werden, ist sehr zeitaufwendig. Das persönliche Risiko besteht darin, dass nicht das ganze Wissen aus den Daten ausgeschöpft wurde und andere die Früchte aus der eigenen Forschung ernten. Aus diesem Grund werden von Forschenden ungern Daten weitergegeben, bei denen sie sich nicht hundertprozentig sicher sind, dass alles relevante Wissen aus ihren Daten ausgeschöpft ist. Das ist global betrachtet ganz offensichtlich suboptimal und ein deutliches Indiz, dass das Anreizsystem zur Datenveröffentlichung unzureichend ist.

Die Abb. 11 zeigt die Häufigkeiten der fehlenden Komponenten, um Forschungsdaten öffentlich zur Verfügung zu stellen. Am häufigsten wird die fehlende Hilfestellung bei der Dokumentation und dem Zugang zu einem Repositoryum genannt. Weitere Antworten waren: Personalmangel und -fluktuation, die Notwendigkeit eines Kulturwandels und fehlende Rechtssicherheit. Hierbei betrifft Rechtssicherheit nicht nur Punkte wie Datenschutz oder Urheberrecht, die in Kapitel 2.1. [Urheberrecht und Datenschutz](#) behandelt werden, sondern auch Pflichten der Forschenden gegenüber den Förderorganisationen.

2.3.2.1 Anreizsystem (83)

„Als Forschende, die eigene Primärdaten gerne publizieren möchten sowie ein besseres Angebot für offen zugängliche Datenbestände für wichtig halten, wünschen wir uns ein Belohnungssystem in Form von Creditpoints für die Publikation von gut strukturierten Daten, um Datenpublikationen für die wissenschaftliche Community attraktiver zu machen.“

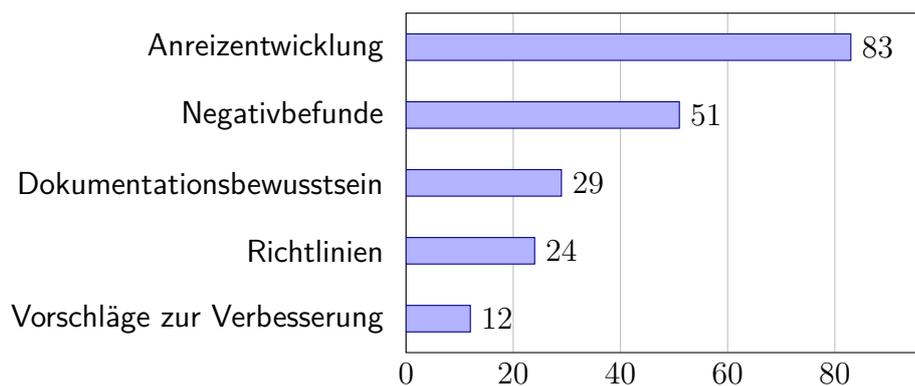


Abb. 12: Angesprochene Themen der Datenkultur

Aus der Entwicklung von Anreizsystemen erhoffen sich Forschende eine bessere Dokumentation und Nachvollziehbarkeit wissenschaftlicher Studien. Die Veröffentlichung von detaillierten Informationen zu Methodenaufbau und Datengewinnung bzw. Reinheit und Ausbeute in Laboren, ist beispielsweise eine Notwendigkeit für die Replikation und Nachvollziehbarkeit von wissenschaftlichen Untersuchungen und somit eigentlich ein Grundpfeiler guter wissenschaftlichen Praxis. Solche User-Stories zum Anreizsystem machen etwa 42 % der Forderungen zur wissenschaftlichen Datenkultur aus. Zwei Drittel dieser Stimmen fordern ganz ausdrücklich eine Verbesserung des Anreizsystems, wohingegen die weiteren zwar in die gleiche Richtung zeigen, aber ohne konkreten Wegvorstellung das Ziel umschreiben, dass sich die Bereitstellungskultur der Forschenden bessern müsse. In der Regel werden Anreizsysteme beschrieben, wie die Berücksichtigung der Datenveröffentlichungen bei der Vergabe von Projektmitteln oder Berufung auf Professuren. Beide Möglichkeiten wurden von Forschenden mehrfach explizit gewünscht.

2.3.2.2 Negativbefunde (51)

„Als Forschende, deren Disziplin durch „Versuch und Irrtum“ voranschreitet, wünschen wir uns dass auch Negativbefunde in einer für uns angemessenen Form zur Verfügung stehen, um diese (anderen schon bekannten) „Irrtümer“ auch in unsere Forschung einfließen zu lassen.“

Unter der Weitergabe von Wissen verstehen die Forschenden auch die Veröffentlichung von negativen Ergebnissen. Mit negativen Ergebnissen sind in diesem Kontext Sackgassen oder Befunde gemeint, die nicht zum gewünschten Ziel geführt haben. In vielen Interviews wurde die Notwendigkeit geäußert, solche Ergebnisse zu veröffentlichen.

Es haben 260 von 578 Interviewten (45 %) die Veröffentlichung von negativen Ergebnissen befürwortet und nur 143 (24,7 %) finden das Konzept nicht sinnvoll. Es ist nicht überraschend, dass sich insbesondere in den Natur- und Lebenswissenschaften nach negativen Ergebnissen gesehnt wird (146 bzw. 52,3 %) und im Bereich der Geistes- und Sozialwissenschaften nur jeder Dritte eine solche Veröffentlichung für sinnvoll hält (57 bzw. 33,7 %). Diese niedrigere Quote bei den Geistes- und Sozialwissenschaften wurde von den Befragten oft dadurch erklärt, dass es sich in diesen Fachbereichen teilweise um qualitative Forschung handelt und deswegen der Anteil von negativen Ergebnissen eher gering ist bzw. so nicht existiert. Ein interessantes

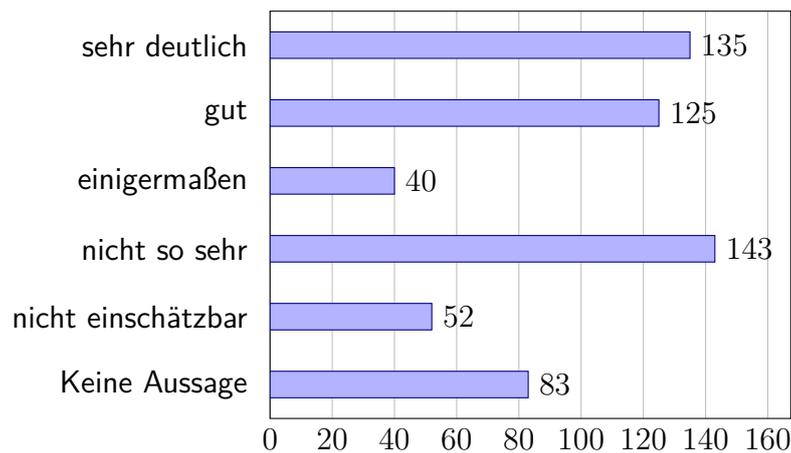


Abb. 13: „Sackgassen“ oder "negative Befunde" werden selten publiziert, weshalb andere Forscher sie oft mühsam reproduzieren müssen. Wie sehr könnten Sie in Ihrem Gebiet von einem Überblick zu solchen Erfahrungen profitieren?“ (Frage 23)

Bild ergibt sich, wenn zu den Antworten die dazugehörigen Bemerkungen der Forschenden näher analysiert werden.

Nicht sinnvoll Viele derjenigen Forschenden, die eine Veröffentlichung von negativen Ergebnissen nicht für sinnvoll erachten, sind der Meinung, dass diese Art von Artikeln nicht relevant für ihre Forschung sind (53,5%) oder es sehr schwierig ist, wissenschaftlich fundiert ein negatives Ergebnis zu beweisen (15,5%). Zudem wurde von einigen Forschenden erwähnt, dass eine entsprechende Anerkennung von Seiten der Verlage und anderer Forschenden fehlt (6,9%).

Sehr deutlich Viele Forschende, die einen Vorteil in der Veröffentlichung von negativen Daten sehen, haben die o.g. Probleme ebenfalls erwähnt, sehen jedoch einen größeren Vorteil in der Veröffentlichung als in der Zurückhaltung von negativen Ergebnissen:

- **Ideenfindung vorantreiben**
Auch wenn Ansätze nicht funktionieren, können diese andere Forschende inspirieren und auf neue Ideen bringen. Durch die Veröffentlichung von fehlgeschlagenen Experimenten wird Forschenden die Angst genommen Fehler zu machen oder zu besprechen.
- **Nachwuchs fördern**
Durch die Wiederverwendbarkeit früherer Arbeiten und die höhere Nachvollziehbarkeit von (negativen) Ergebnisse haben junge Forschende einen leichteren Einstieg in die Wissenschaft und können auf den Erfolgen (und Fehlern) von vorhergehenden Forschenden aufbauen.
- **Effizienz steigern**
Durch die Vermeidung von „Sackgassen“ kann die zur Verfügung stehende Zeit besser

verwendet und neue Ansätze untersucht werden. So findet keine unnötige Doppelfinanzierung von wissenschaftlichen Studien statt. Manche Studien werden in einzelnen wissenschaftlichen Disziplinen immer wieder durchgeführt, weil kaum Informationen vorhanden sind, welche Experimente und wissenschaftlichen Studien bereits existieren. Die Finanzierung von Untersuchungen, die bereits in der Vergangenheit fehlgeschlagen sind, kann verhindert werden, indem diese regelmäßig veröffentlicht werden und Forschende sowie Gutachter sich darüber informieren können.

2.3.2.3 Richtlinien (24)

„Als Forschende, wünschen wir uns praxistaugliche Regeln und Richtlinien zum Datenumgang, um die eigenen Datenverwaltungsprozesse daran auszurichten und von anderen Gleiches fordern zu können.“

Viele Forschende wünschen sich von Förderorganisationen und Behörden neben einem Anreizsystem auch klare Vorgaben für die Forschung bzgl. der Veröffentlichung von Daten. Diese Richtlinien sollen durch Datenqualitäts- und Granularitätskriterien einen direkten Einfluss auf die Datenveröffentlichungen ausüben und stellen zum einen das Gegenstück zur gewünschten „Möhre“ [2.3.2.1. Anreizsystem](#), zum anderen aber auch eine konkrete Orientierungshilfe zur Verfügung. So verlangen die Forschenden beispielsweise von den Förderorganisationen klare Richtlinien bzgl. der Aufbewahrungspflicht und -art von Daten sowie die Veröffentlichung von negativen Ergebnissen. Es wurde häufig der Wunsch ausgedrückt, von Auflagenveröffentlichungen wegzukommen und bei der Veröffentlichung auf die wissenschaftliche Qualität und Signifikanz mehr Wert zu legen.

2.3.2.4 Dokumentationsbewusstsein (29)

„Als Forschende, die Widerlegbarkeit und Reproduzierbarkeit in ihrer Disziplin als vorrangig erachten, wünschen wir uns, dass die wesentlichen (auch kleinen) Operationen eines Forschungsvorgangs sowie Datenbeschreibungen vollständig veröffentlicht werden, um andere Forschungsarbeiten seriös bewerten zu können.“

Die Forschenden sehen sich hier auch selbst und ihre Mitforschenden in der Bringschuld und fordern bei wissenschaftlichen Arbeiten ein Mindestmaß an Transparenz, sodass diese reproduziert und ggf. widerlegt werden können. Es geht bei der Reproduzierbarkeit darum, auch Forschungsdetails aufzuführen, sodass ein Ergebnis nicht nur grob nachvollziehbar, sondern tatsächlich auch im Detail reproduzierbar ist. Dass die dafür nötige Dokumentation oft nicht vorhanden oder von zu niedriger Qualität ist, wird in diesen User-Stories bemängelt. In einigen Fällen werden die Ergebnisse anderer Forscher auch angezweifelt, ohne dass man jedoch in der Lage wäre die Resultate (aufgrund der schlechten Dokumentation/Reproduzierbarkeit) zu widerlegen.

2.3.2.5 Vorschläge zur Verbesserung eines Teilaspektes (12)

Die restlichen User-Stories nehmen oft konkret Bezug auf einen bestimmten Aspekt der Datenkultur und machen auch Vorschläge zur Verbesserung. Beispielsweise wird vorgeschlagen,

dass Gutachter immer Zugang zu allen Daten und Programmen erhalten sollten oder dass die Daten zu allen Schaubildern in Papern gleich in diese integriert werden sollten. Auch dass es nach Konferenzen ein Standardverfahren geben sollte, wie Daten (zu Postern und Vorträgen) zur Verfügung gestellt werden, wurde hier genannt.

2.3.3 Status quo

Die Forschenden sind mit der aktuellen Lage des Datenaustauschs zu über einem Drittel unzufrieden. Es findet zwar ein regelmäßiger Austausch zwischen den Forschenden statt, allerdings nicht in dem Ausmaß, wie es auch von den Forschenden selbst gewünscht ist. In Frage 20 haben wir erfasst, in welchem Rahmen die Forschenden bereits Primärdaten ausgetauscht haben. Obwohl fast die Hälfte der Befragten ausdrücklich erwähnt hat, dass ihre Daten für andere interessant sein könnten, teilt ein Großteil der Forschenden Daten nur innerhalb des Projekts und / oder innerhalb der Forschungsgruppe. Nur ungefähr 10% der Forschenden stellen solche Daten öffentlich zum Download bereit. Das Balkendiagramm in [Abb. 14](#)) gibt einen Überblick die aktuelle Situation.

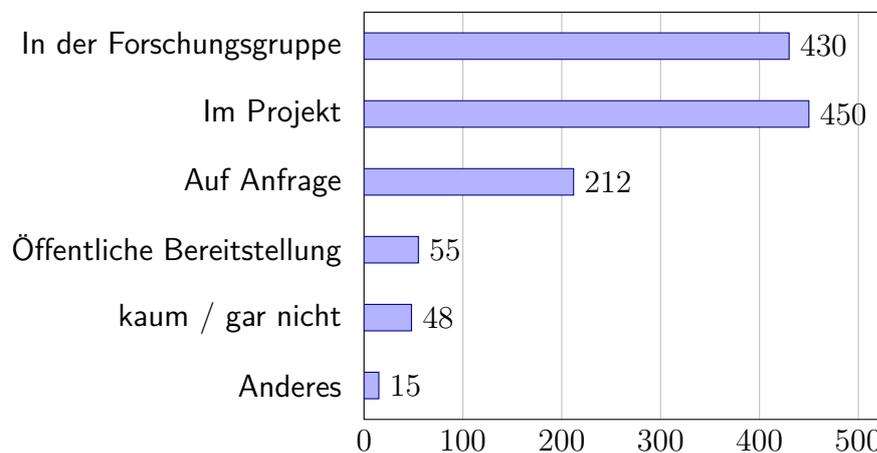


Abb. 14: „In welchem Rahmen haben Sie schon eigene Primärdaten ausgetauscht?“
(Frage 20a – Mehrfachnennung der vorgegebenen Austauschvarianten möglich)

Wie in allen Bereichen der Leistungserbringung durch unterschiedliche Parteien herrscht auch in der Wissenschaft eine Konkurrenzsituation. Jedes Jahr werden unzählige wissenschaftliche Artikel veröffentlicht. Auf dem aktuellen Stand der Technik zu bleiben wird für den einzelnen Forschenden immer schwieriger. Aus diesem Grund ist es wichtig, die wissenschaftlichen Arbeiten und somit auch Ihre Verfasser zu bewerten und eine Auswahl der lesenswerten Artikel zu treffen.

Im Laufe der Zeit sind verschiedene Methoden zur Beurteilung von Forschenden entstanden: h-index, i10-index und g-index.

Der h-index gibt die Anzahl der Artikel pro Person an, die h Artikel mit mindestens h Zitaten veröffentlicht hat. Der i10-index gibt die Anzahl der Artikel an, die mindestens zehnmal zitiert wurden. Der g-index ist die Anzahl der am häufigsten zitierten Artikel, die akkumuliert mindestens g^2 Zitierungen aufweisen.

Alle Methoden zur Bewertung von Forschenden beruhen auf der Anzahl der Zitationen, die seine Artikel erhalten. Die Grundprämisse bei dieser Vorgehensweise ist die folgende: Je häufiger ein Artikel von anderen Forschenden zitiert wird, desto wichtiger ist der Artikel. Diese Methode der Bewertung ermöglicht eine einfache und quantitative Bewertung über die Wichtigkeit eines Artikels. Jedoch führt sie zu einer wissenschaftlichen Gesellschaft in der Forschende nur nach ihren zitierten Artikeln bewertet werden.

Die Benutzung von Zitaten als Qualitätsmerkmal kann zu einer Fokussierung auf bestimmte Themengebiete führen. Zum einen konzentrieren sich Verlage auf „Sensationsmeldungen“ und beurteilen Artikel anhand der potentiellen Publizität, die sie erzeugen und zum anderen forschen Forschende in den aktuell populären Themen, weil hier eine Veröffentlichung und damit die Anerkennung sicher zu sein scheint.

Genauso wie Forschende, werden auch Verlage bzw. Journale anhand der von ihnen veröffentlichten und zitierten Artikel gemessen. Eine Übersicht des aktuellen h5-index (h-index bezogen auf die letzten fünf Jahre) von Journalen kann unter Google Scholar³⁶ betrachtet werden.

2.3.3.1 Förderorganisationen und ihre Stellung zu Forschungsdaten (DE)

Deutsche Forschungsgemeinschaft (DFG) Im „Leitfaden für die Antragstellung“³⁷ wird in Kap. 2.4 auf den Umgang mit den im Projekt erzielten Forschungsdaten eingegangen. Hier werden Forschende aufgefordert anzugeben „ob und auf welche Weise [die Forschungsdaten] für andere zur Verfügung gestellt werden“. Die Gemeinschaft unterstützt die Aufbereitung der Forschungsdaten für die Nachnutzung, indem die anfallenden projektspezifischen Kosten beantragt werden können.

Bundesministerium für Bildung und Forschung (BMBF) Das BMBF unterstützt seit Oktober 2013 in seinem Verbund „Forschungsdaten Bildung“³⁸ die Sicherung von Daten aus der empirischen Bildungsforschung. Soweit möglich sollen diese Daten außerdem für die Nachnutzung verfügbar gemacht werden. Man muss hier deutlich sagen, dass die Forschenden das aktuelle Angebot, um ausdrücklich Fördermittel für das FDM einzuwerben oft nicht ausreizen.

2.3.3.2 Förderorganisationen und ihre Stellung zu Forschungsdaten (Ausland)

Die Forschungseinrichtungen im Ausland haben die Wichtigkeit des Themas Forschungsdatenmanagement früh erkannt und geben teilweise schon eindeutige Vorschriften an die Forschenden. Viele Förderorganisationen haben strikte Regeln zur Datennutzbarmachung und Vorgehensweisen definiert, an die sich die geförderte Institution halten muss.

³⁶ https://scholar.google.de/citations?view_op=top_venues&hl=de [letzter Zugriff 13.5.2015].

³⁷ 54.01 – Leitfaden für die Antragstellung – Projektanträge, http://www.dfg.de/formulare/54_01/ [letzter Zugriff 13.5.2015].

³⁸ http://www.bmbf.de/pub/Flyer_Verbund_Forschungsdaten_Bildung.pdf [letzter Zugriff 13.5.2015].

Großbritannien Der Nationale Forschungsrat Großbritanniens (RCUK) veröffentlichte im Juni 2005 eine Stellungnahme³⁹ mit Bezug auf Forschungsergebnisse. In dieser Stellungnahme heißt es: „RCUK also notes that one of the benefits of digitisation and publication in digital formats is the ability to provide access to primary research data alongside the traditional article; and it shares the Select Committee’s and the Government’s view that the data underpinning the published results of publicly-funded research should be made available as widely and rapidly as possible.“ Für die Veröffentlichung der Forschungsdaten stellen zwei der Mitgliedsorganisationen (ESRC und NERC) Dienstleistungen zur Bestandserhaltung und weitere zwei Organisationen (BBSRC und MRC) akzeptieren die Langzeitkosten zur Erhaltung der Daten beim ursprünglichen Antrag. Die übrigen Institutionen (EPSRC, AHRC und STFC) haben Vereinbarungen mit Rechenzentren um mindestens teilweise Daten sichern zu können.

Vereinigte Staaten von Amerika Die zwei größten Forschungsförderorganisationen (NIH und NSF) in den USA haben 2005 bzw. 2010 die Wichtigkeit von Forschungsdaten erkannt und erste Regeln zum Umgang mit diesen Daten aufgestellt.

Das National Institute of Health (NIH) verlangt seit Mai 2005 von allen Forschenden, die eine wissenschaftliche Arbeit aus einem NIH (teil-)finanzierten Projekt veröffentlichen, dass dieser Artikel zusätzlich ins öffentlich zugängliche PubMed Central der National Library of Medicine (NLM) gestellt wird. Zusätzlich muss bei der Beantragung einer NIH-Förderung ein Data-Sharing-Plan hinzugefügt werden. Falls eine Veröffentlichung nicht vorgesehen ist, muss erklärt werden, wieso die Veröffentlichung der Daten nicht möglich ist.

Seit 2010 fordert die National Science Foundation (NSF) in den USA, dass die von Ihnen geförderten Projekte in Zukunft sogenannte Datenmanagementpläne aufweisen müssen⁴⁰. In Abschnitt IV.4.b. vom Grand Proposal Guide (GPG)⁴¹ des NSF heißt es: „Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants.“

Viele weitere Förderorganisationen (wie z. B. das United States Department of Agriculture, National Aeronautics and Space Administration und United States Agency for International Development) sind dem Beispiel von NIH und NSF gefolgt und fordern Datenmanagementpläne bei der Beantragung von Fördergeldern.

OECD Die Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) veröffentlichte 2004 die „Declaration on Access to Research Data from Public Funding“⁴² für einen einheitlichen Zugang zu Forschungsdaten aus öffentlichen Mitteln nach Prinzipien der Offenheit, Flexibilität, Transparenz und Interoperabilität. Empfehlungen der OECD sind

³⁹ <http://www.rcuk.ac.uk/RCUK-prod/assets/documents/documents/2005statement.pdf> [letzter Zugriff 13.5.2015].

⁴⁰ http://nsf.gov/news/news_summ.jsp?cntn_id=116928&org=NSF [letzter Zugriff 13.5.2015].

⁴¹ http://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/gpg_index.jsp [letzter Zugriff 13.5.2015].

⁴² <http://www.unitus.it/biblioteche/webif/06-docs/download/openaccess/oecddeclaration.pdf> [letzter Zugriff 13.5.2015].

zwar keine rechtlichen Anordnungen, jedoch haben sie einen gewissen Leitliniencharakter, nach denen sich die Mitgliedsländer richten und sind deshalb ein wichtiges Instrument zur Aufklärung über Probleme.

EU-Rahmenprogramm Horizon 2020 Das Rahmenprogramm Horizon 2020 wird von der EU als Pilotprogramm für „Open Research Data“⁴³ verwendet. Die Forschenden müssen Ihre Forschungsdaten, auf denen die veröffentlichten Artikel beruhen, der Öffentlichkeit zur Verfügung stellen. Damit hält die Europäische Kommission ihr Versprechen, das sie im Juli 2012 (siehe IP-12-790⁴⁴) gegeben hat, dass der Zugang zu wissenschaftlichen Daten verbessert wird.

2.3.4 Gegenwärtige Handlungsempfehlungen

Es ist schwierig Empfehlungen zu definieren, die eine langfristige Änderung in der Einstellung gegenüber der Offenheit von Forschungsdaten bewirkt. Jedoch könnten verschiedene parallele Ansätze hilfreich sein, die Wertschätzung von Datenpublikationen zu erhöhen.

2.3.4.1 Anreizsystem

Das Wissenschaftssystem müsste so gestaltet sein, dass diejenigen, deren veröffentlichte Daten einen Effekt zuerst erkennen lassen, ebenso zu honorieren sind, wie diejenigen, die die Beobachtung des Effektes später machen. Es kann durchaus der Prozess der Datenerzeugung die größere Leistung sein, insbesondere in Fachgebieten, in denen die hochwertige Datenerzeugung schwierig ist. Die saubere Dokumentation käme in so einem Umfeld automatisch, denn nur diese ermöglicht, dass andere mit den Daten (zum eigenen Vorteil) arbeiten können. Es ist gerade ein Zeichen einer sauberen wissenschaftlichen Leistung, wenn andere, in ihnen eigentlich fremden Datensätzen, bemerkenswerte Dinge entdecken können. Eine gewisse Datenqualität sollte natürlich dennoch die Voraussetzung für die Leistungsanerkennung sein. Dies könnten Repositorien sicherstellen.

2.3.4.2 Unterstützung durch Förderorganisationen und Universitäten

Es ist wichtig, die Verantwortung für gutes Forschungsdatenmanagement nicht nur der kleinsten Einheit (dem einzelnen Forschenden) anzulasten, sondern auf höherer Ebene Stellung zu beziehen. Falls eine Änderung der Datenkultur nicht von höherer Instanz angegangen wird, besteht das Risiko, dass sich Insellösungen bilden. Dies würde eher demotivierend für den einzelnen Forschenden wirken und dazu führen, dass Forschende sich mit den damit verbundenen Problemen allein gelassen fühlen. Weiterhin könnten Datenpublikationen stärker als Kriterium für die Berufung von Professoren genutzt werden.

⁴³ http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf [letzter Zugriff 13.5.2015].

⁴⁴ http://europa.eu/rapid/press-release_IP-12-790_en.htm [letzter Zugriff 15.6.2015].

2.3.4.3 Aufklärung zur Unterstützung durch Förderorganisationen

Hier Bedarf es der Aufklärung über die Möglichkeiten und vielleicht auch einer Prüfung durch Forschungsförderungsabteilungen an den Universitäten, die darauf hinweisen in Projekten immer auch Mittel für FDM zu beantragen. In DMP-Instrumenten ist ein solcher Punkt leicht zu verankern.

2.3.4.4 Hinweise für Forschende

Die Anforderung vieler Forschender an sich und andere ist es, Transparenz zu zeigen und Arbeiten nachvollziehbar zu veröffentlichen, sodass diese reproduziert werden können. Die Nachvollziehbarkeit der Ergebnisse kann in einigen Fällen recht einfach durch die Benutzung von SVN / GitHub-ähnlichen Systemen für die Abspeicherung von Modellen, Problembeschreibungen, Eingangsdaten und Programmcode erreicht werden. Solche Empfehlungen systematisch zu etablieren und konkrete Richtlinien auszuarbeiten ist Aufgabe der Fachbereiche. Allerdings kann eine zentrale Anlaufstelle für Datenmanagement mit Best Practices und Hilfen für die Forschenden einen zusätzlichen Beitrag leisten.

Abschließend lässt sich zusammenfassen, dass die meisten Akteure eine Modernisierung des Publikationssystems befürworten, doch noch zu wenig echte Alternativen zum bestehenden System existieren. Mit unseren Vorschlägen in diesem und anderen Kapiteln hoffen wir die dringend notwendige Diskussion auf Basis konkreter Ideen weiter vorantreiben zu können.

3 Digitale Forschungsdatenverarbeitung

3.1 Forschungsnahe Unterstützung (271)

3.1.1 Einleitung

Dieses Unterkapitel beschäftigt sich mit 271 User-Stories mit Bezug zur Verarbeitungs von Forschungsdaten, in denen ein Bedarf nach Unterstützung geäußert wird. Allgemein zusammengefasst sind für übliche IT-Aufgaben, die mit der Forschungsarbeit einhergehen, durch den digitalen Wandel immer mehr Kompetenzen erforderlich, die von der üblichen Fachausbildung abweichen. Zudem folgt aus der vermehrten Arbeit mit Forschungsdaten die Notwendigkeit, sich mit FDM auseinanderzusetzen, um den vollen Nutzen aus der investierten Arbeit zu ziehen und langfristig erfolgreich zu sein. Daraus ergibt sich eine Nachfrage nach Hilfsangeboten, die fehlende Kompetenzen durch Beratungen oder Übernahme von Aufgaben ausgleichen sollen. Es ist eine leichte Häufung in den Geistes- und Sozialwissenschaften und speziell in den Digital Humanities festzustellen, wo zunehmend Aufgaben mit Programmieranforderungen anfallen, die bisher nicht Teil des Kurrikulums in diesen Wissenschaften sind (s. [Abb. 15](#)).

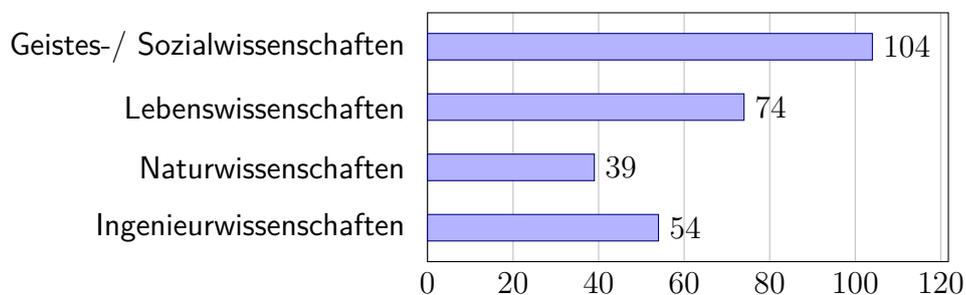


Abb. 15: Verteilung der User-Stories nach der Fachsystematik der DFG

Ein wichtiger Teilbereich sind solche Aufgaben, deren Durchführung ausgelagert oder in Kooperation mit zentralen Stellen geleistet werden kann. Es gibt eine gewisse Verwandtschaft mit [2.2. Informationsangebote](#), denn der Übergang zwischen IT-Beratung und aktiver Unterstützung ist fließend, etwa wenn Beratung zur Umsetzung technischer Details gesucht ist. Insgesamt gibt es 15 dieser Grenzfälle in den User-Stories, die in beide Kapitel einfließen. User-Stories die nicht im engeren Sinne forschungsnahe sind, sondern Unterstützungswünsche mit Bezug zu alltäglicher EDV und Systemadministration werden nicht hier sondern ausschließlich in [6.2. IT-Support](#) behandelt. Hier gibt es aber ebenfalls 13 Überschneidungen.

Weitere 12 Überschneidungen gibt es mit Projekt- und Datendokumentation. 11 User-Stories sind zwar forschungsnah, aber so nah, dass die zugrundeliegende Unterstützungsforderung von der eigentlichen wissenschaftlichen Arbeit kaum zu unterscheiden ist und werden deshalb im folgenden Abschnitt nicht weiter beachtet.

3.1.2 Bedarf der Wissenschaftsgemeinschaft

Im Verlauf der geführten Interviews wurde versucht die Frage zu beantworten, was den Forschenden bei ihrer Arbeit im Umgang mit Daten im Weg steht. Die Befragten berichteten häufig, dass ein großer Teil der Arbeitszeit aufgewendet wird, um Daten für das jeweilige Feld zu erheben, aus externen Quellen zusammenzustellen und letztendlich verwendbar zu machen oder sich mit alltäglicheren technischen Anforderungen auseinanderzusetzen.

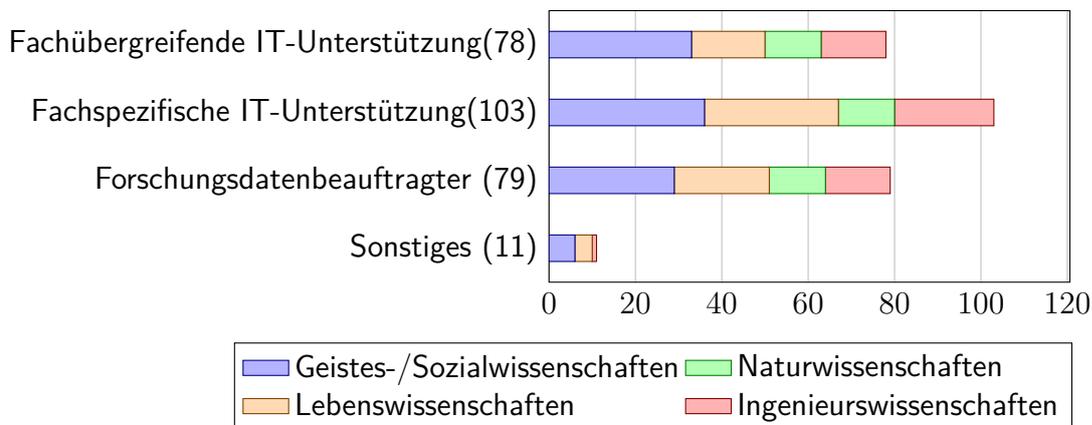


Abb. 16: Verteilung der Unterthemen nach der Fachsystematik der DFG

Die verschiedenen fachspezifischen Aufgaben, bei denen IT-Fachwissen fehlt, oder aber bei denen der Wunsch einer Begleitung der Datenverwaltung im Vordergrund steht, sind im Folgenden beschrieben. Die Verteilung der anfordernden Wissenschaftsbereiche innerhalb dieser drei Gruppen ist relativ unauffällig. Insgesamt

3.1.2.1 Fachübergreifende forschungsnaher IT-Unterstützung (78)

„Als Forschende, wünschen wir uns eine Anlaufstelle, um bei Datenerhebung oder Datenverarbeitung Unterstützung durch Ausleihkräfte bekommen zu können.“

Die Wünsche nach allgemeiner IT-Unterstützung dienen in der Regel dazu, Programmieraufgaben zu lösen, die Forschende selbst nicht effizient bewältigen können. Die Trennlinie zu fachspezifischer IT-Unterstützung ist natürlich fließend und wurde in der Regel⁴⁵ so gezogen, dass fachspezifische Schnittstellen zu kennen noch der allgemeinen IT-Unterstützung zugeordnet ist, wohingegen in fachspezifischer Software selbst zu arbeiten als fachspezifisch klassifiziert wurde. Ausnahme ist hier die Hilfe zur Parallelisierung, die ebenfalls als fachübergreifend festgelegt ist. Das fachübergreifende Aufgabenfeld ist sehr heterogen und kann

⁴⁵ Nicht, wenn der Befragte explizit nach IT-Fachkräften mit disziplinspezifischen Wissen verlangt hat.

hier nur umrissen werden. Typische Aufgaben umfassen (einfache) automatisierte Bildanalyse, Web-Crawling, Softwareberatung, HPC-Optimierung und Parallelisierung, Datenkompression, Formatkonvertierung sowie allgemeine Unterstützung bei anderen spezifischen Programmierproblemen.

3.1.2.2 Fachspezifische IT-Unterstützung (103)

„Als Forschende mit hohem IT-Bedarf, wünschen wir uns eine fachspezifische Anlaufstelle, um Beratung zu bekommen und Unterstützung bei der Entwicklung von IT-Lösungen für unsere Forschung in Auftrag geben zu können.“

Die fachspezifischen IT-Anforderungen kommen aus den Bereichen Softwareentwicklung, Hilfe bei statistischer Analyse und IT-Beratung.

Manche Forschende wollen die Softwareentwicklung komplett outsourcen, andere benötigen eine Software zur Versuchssteuerung oder wollen nur die Performance ihrer Programme verbessern. Die fachspezifische IT-Beratung umfasst bspw. Fragen zur Benutzung/Empfehlung von Fachsoftware oder eine Machbarkeitsanalyse einer gewünschten Softwareimplementierung. Die statistische Analyse von Daten benötigt, wie alle fachspezifischen Aufgaben, eine gewisse Einarbeitungszeit einer IT-Fachkraft in das jeweilige Themengebiet. In der Regel würden die Aufgaben wohl bei zufriedenstellender Ausführung von den Forschenden auch bezahlt werden. Geld ist hier nicht das Hauptproblem, sondern der Mangel an Expertise in der Forschungsgruppe, der viel mehr kostet, weil dadurch ineffiziente Programmentwicklung durch IT-Laien betrieben werden muss.

3.1.2.3 Verantwortliche für Forschungsdaten (79)

„Als Forschende, wünschen wir uns eine verantwortliche Stelle, um uns bei FDM aktiv zu unterstützen.“

Verantwortliche für Forschungsdaten (VFD) wären Personen, die an einzelnen Instituten, Institutsverbänden oder zentral angesiedelt sein könnten, deren Arbeitsschwerpunkt aber disziplinspezifisch ist. Die User-Stories befassen sich nicht mit der Aufbereitung von Daten zur Publikation, sondern mit einer konzeptuellen Hilfe, die dauerhaft zur Verfügung steht. Mehrheitlich wird sich eine Fachkraft gewünscht, die die FDM-Forschungsinfrastruktur (Datenbanken, Repositorien, Quellcodes) bereitstellt, pflegt oder aufsetzt. In den restlichen User-Stories wäre eine Person erwünscht, die die FDM-Standards der Gruppe überwacht, vereinheitlicht und die Gruppenleistungen langfristig verfügbar hält.

3.1.3 Status quo

Informationen zum allgemeinen IT-Support an den Universitäten enthält [6.2.3. Status quo](#). Was die spezialisierten Unterstützungsangebote betrifft, so gibt es vor allem in den Lebenswissenschaften⁴⁶, an einigen Standorten Core-Facilities, die die Forschenden bei wiederkehrenden Analyseaufgaben unterstützen. Außerdem gibt es Institute, die ihre Expertise an

⁴⁶ Bspw. http://www.embl.de/services/core_facilities, [letzter Zugriff 30.6.2015]

ihrer Universität (bspw. in der Visualisierung) einbringen. Im HPC-Bereich wird der Ausbau von Supportteams in Baden-Württemberg vorangetrieben. Trotz all dieser vereinzelter Maßnahmen sind im Allgemeinen fachspezifische Unterstützungsangebote noch nicht für die Forschenden vorhanden. Auch die ITBs der Institute kennen die IT manchmal nur auf dem Papier und sind an kleinen Instituten oft ganz normale Forschende, die für ihre Arbeit als ITB keine zusätzlichen Ressourcen haben. Der allgemeine IT-Support kann die in diesem Kapitel zusammengetragenen Bedarfe nicht erfüllen. In Mannheim gibt es ein Forschungsdaten Service-Center⁴⁷, das forschungsnaher IT-Bedarfe deckt und sich dabei auf einen Mitarbeiter und mehrere wissenschaftliche Hilfskräfte stützt.

3.1.4 Gegenwärtige Handlungsempfehlungen

Die drei Empfehlungen des Kapitels scheinen jeweils zu den einzelnen Unterthemen zu passen, sind jedoch nicht als Alternativen oder disjunkte Ansätze, sondern lediglich als Möglichkeiten zu verstehen. Jegliche Kombination der Ansätze und verschiedenste Implementierungen scheinen machbar.

Science@home Für kleine IT-Aufgaben, die einfach zu spezifizieren und zu prüfen (aber womöglich schwer zu lösen) sind, könnte ein gut sichtbares E-Science-Zentrum versuchen eine Community aus Freiwilligen aufzubauen, die die Wissenschaft analog zu den Projekten des Verteilten Rechnens⁴⁸ unterstützen. Ist der Ansatz erfolgreich, könnte mit wenigen bezahlten Stellen viel Wissenschaftsunterstützung möglich werden. Hier sollten die Landesuniversitäten ein gemeinsames Projekt anstreben.

Ausleihservice am Kompetenzzentrum Eine spezifischere Möglichkeit auch komplexere Zusammenarbeit mit den Forschenden zu ermöglichen, wäre die Einrichtung eines Pools aus IT-Kräften, die die Forschenden projektbegleitend oder in kurzen Phasen unterstützen. Die Universitäten können bei der Ausrichtung eines solchen Pools leicht eigene Schwerpunkte setzen. Hier ist es wichtig, die ganze Palette aus kurzen Aufträgen und langer intensiver Zusammenarbeit abbilden zu können und dafür ein praktikables Kostenmodell zu haben. Ob solche Zentren von ITBs, dem allgemeinen IT-Service oder direkt von Forschenden ansprechbar sein sollten, hängt von der genauen Einbettung in das Feld dieser anderen Akteure ab.

Verantwortliche für Forschungsdaten Die Verantwortung für FDM liegt letztlich bei den Forschenden selbst. Doch der Bedarf nach klaren Strukturen, um Standards einzuhalten und einen effizienten Umgang mit Daten zu etablieren, kann nicht auf der Ebene des einzelnen Forschenden gelöst werden. VFDs, die Repositorien bereitstellen, Datenbanken pflegen und darauf achten, dass Daten korrekt abgelegt werden, müssen den Spagat aus Gängelung und Hilfestellung schaffen können, um alle Forschenden auf einen zumindest lokal standardisierten Stand zu bringen. Hier könnte klein angefangen oder sich bemüht werden im Rahmen von

⁴⁷ <http://service.informatik.uni-mannheim.de/> [letzter Zugriff 30.6.2015]

⁴⁸ https://de.wikipedia.org/wiki/Liste_der_Projekte_verteilten_Rechnens - [letzter Zugriff 30.6.2015].

SFBs solche Stellen zu etablieren. Langfristig könnten die Universitäten anstreben, dass ein kleiner Anteil des wissenschaftlichen Personals, auf solche Aufgaben spezialisiert ist. Auch ein Bottom-up-Szenario ist denkbar, in dem Forschende selbst diese Aufgaben übernehmen (was sie aktuell ja tun). Allerdings sorgt hier das wissenschaftliche Anreizsystem dafür, dass solche Aufgaben nicht sehr attraktiv sind. Vielleicht könnte hier eine bessere Anerkennung von Datenveröffentlichungen, die VFBs gemeinsam mit den Forschenden schreiben dafür sorgen, dass so ein Szenario realisierbar wird. Auch eine Besetzung mit Dauerstellen wäre denkbar.

3.2 Wissenschaftliche Softwaretools (157)

3.2.1 Einleitung

Moderne Software in Kombination mit steigender Rechenleistung und wachsenden Speicherkapazitäten ermöglicht die Durchführung von wissenschaftlichen Analysen und Experimenten, die noch vor wenigen Jahrzehnten technisch unmöglich waren. Dazu gehören zum Beispiel automatisierte Bild- und Textanalysen, komplexe statistische oder bioinformatische Auswertungen und Simulationen.

Dieses Kapitel beschreibt die in den Interviews erfassten Bedarfe von Forschenden im Bereich der wissenschaftlichen Softwaretools. Es basiert auf 157 User-Stories. Darunter befinden sich, im Verhältnis zur Stichprobengesamtheit, vor allem viele aus den Ingenieur- und Lebenswissenschaften (vgl. [Abb. 17](#) mit [Abb. 1](#)).

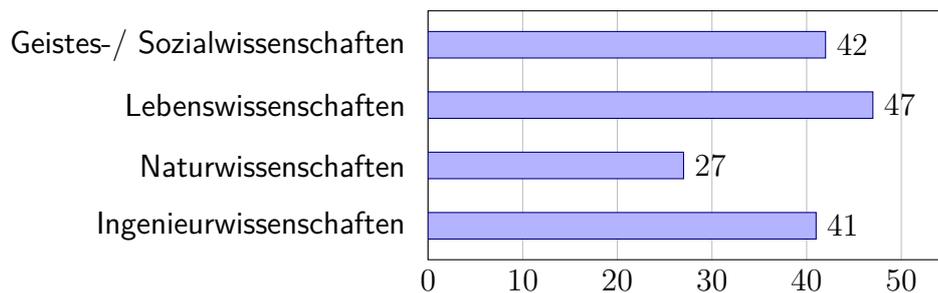


Abb. 17: Verteilung der Unterthemen nach der Fachsystematik der DFG

3.2.2 Bedarf der Wissenschaftsgemeinschaft

Die angesprochenen Bedarfe der Forschenden sind in [Abb. 18](#) zusammengefasst und nach Wissenschaftsbereichen aufgeteilt dargestellt. Je nach Zweck variiert die wissenschaftliche Zusammensetzung einer Bedarfsgemeinschaft.

In diesem Kapitel liegt der Fokus auf den Funktionalitäten, die sich die Forschenden wünschen. Woher die Software kommen soll (Einkauf, Eigenentwicklung, externes Entwicklungsprojekt, ...) wird hier nicht betrachtet.

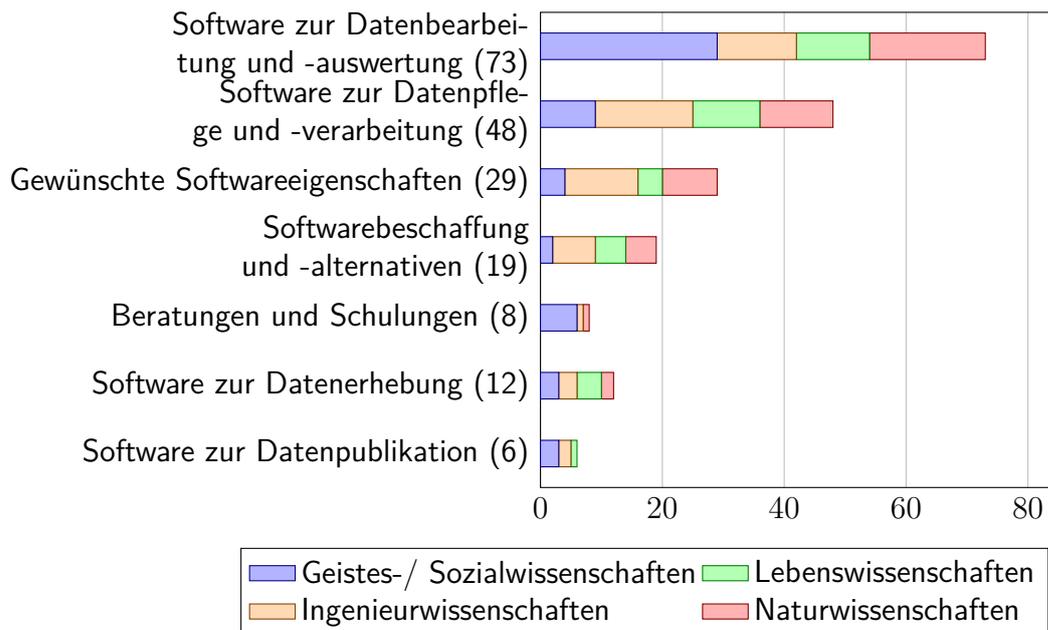


Abb. 18: Verteilung der Unterthemen nach der Fachsystematik der DFG

3.2.2.1 Software zur Datenbearbeitung und Datenauswertung (73)

„Als Forschende, wünschen wir uns Softwarewerkzeuge, um unsere Daten effizienter aufzubereiten und auszuwerten.“

Die Datenverarbeitung stellt den führenden Bedarf innerhalb der Softwaretools dar. Diese Bedarfe umfassen Software zur Arbeit mit unterschiedlichen Datentypen: Bildern, Texten, Audiodateien sowie GIS-Daten. So wurde von den Forschenden der Bedarf nach bestimmter Bildauswertungssoftware sowie nach Softwarelösungen zur Texttranskription, Sprachsegmentierung und zur Textanalyse geäußert. Ebenso wurden spezielle Konverter zur Konvertierung von Zeichensätzen, OCR-Tools sowie Softwareprodukte zur Texteditierung gewünscht.

Im Bereich Geoinformationssysteme (GIS) wäre die Möglichkeit einer Verknüpfung und Analyse von GIS-Daten aus verschiedenen Datenquellen sowie eine anschauliche Visualisierung von geographischen Datenpunkten wünschenswert. Auch eine mobile Applikation für Smartphones zur Eintragung von topologischen Daten während der Durchführung von Feldexperimenten fällt in die Kategorie „Softwarewerkzeug zur Datenbearbeitung“.

Ausgesprochen wurden auch Wünsche nach Software zur Durchführung von statistischen Analysen, Data Mining sowie Simulationen und Modellierungen, Visualisierungen oder zur kollaborativen Analyse von Daten.

3.2.2.2 Software zur Datenpflege und besseren Datenverarbeitungsabläufen (48)

„Als Forschende mit komplexen Daten, wünschen wir uns Softwarewerkzeuge, um die Datenpflege effizienter zu organisieren und Arbeitsabläufe zu optimieren.“

Forschende haben einigen Softwarebedarf zur Datenpflege- sowie zur Ablaufoptimierung geäußert, um ihre Forschungsarbeit effizient ausführen zu können. Die meisten angesprochenen Anforderungen in diesem Themenkreis betrafen Softwarewünsche zur Datenverwaltung und -beschreibung. Das umfasst Wünsche nach angepasster Software für elektronische Laborbücher ebenso wie Datenbanksysteme zur Datenverwaltung oder Konverter zwischen verschiedenen Formaten, sowie spezielle Schnittstellen und Werkzeuge zur automatisierten Aufbereitung von erhobenen Daten. Am Ende des Datenlebenszyklus sind Systeme zur Speicherung und Archivierung von Forschungsdaten gewünscht. Bedarf nach Versionierung von Forschungsdaten sowie deren Signierung zwecks Manipulationsvermeidung wurde von einigen Forschenden ebenso geäußert. Zusätzlich wäre es wünschenswert, wenn die digitalen Werkzeuge auch verstärkt auf Datensicherheit und Datenschutz Wert legen würden.

3.2.2.3 Gewünschte Softwareeigenschaften (29)

„Als Forschende, die einen hohen Wert auf die Funktionalität der eingesetzten Software legen, wünschen wir uns zusätzliche Softwareeigenschaften, um bestimmte Anwendungen in unserer Forschungsarbeit einfacher und effizienter einsetzen zu können.“

Zu gewünschten Softwareeigenschaften gehören den Aussagen zur Benutzerfreundlichkeit, Interoperabilität, hohen Kompatibilität mit anderen für die Forschenden relevanten Softwaretools sowie Nachhaltigkeit des Softwareangebots. Open Source wurde in zehn User-Stories als eine wünschenswerte Softwareform explizit angegeben.

3.2.2.4 Softwarebeschaffung und Softwarealternativen (19)

„Als Forschende, die für ihre Arbeit regelmäßig Software beschaffen müssen, wünschen wir uns Softwarebeschaffung mit bestimmten Eigenschaften bzw. kostenlose Alternativen zu manchen von uns aktuell eingesetzten Anwendungen, um kostengünstig und einfach die für uns wichtigen Softwarelösungen bekommen zu können.“

Im Themenfeld Softwarebeschaffung und Softwarealternativen wurden mehr Mittel für Softwarebeschaffung gewünscht sowie der Bedarf nach kostenlosen offenen Softwarelösungen bzw. Alternativen zu bestimmten Softwareprodukten geäußert. Mehrere Forschende haben den Wunsch nach kostenlosen Alternativen zu Softwareprodukten der Firma Esri⁴⁹ sowie zu MatLab⁵⁰-, SigmaPlot⁵¹- und GraphPad Prism⁵²-Software ausgesprochen. Der Wunsch nach einer zentralen Softwarebeschaffung sowie nach günstigen bzw. kostenlosen Softwarelizenzen wurde ebenso genannt wie mehr Akzeptanz für Open Source. Details mit Aspekten zu diesen Themen finden sich in [6.3. Lizenzierung](#).

⁴⁹ Homepage von Esri www.esri.de [letzter Zugriff 30.6.2015].

⁵⁰ Homepage von Matlab <http://de.mathworks.com/products/matlab/> [letzter Zugriff 30.6.2015].

⁵¹ Homepage von SigmaPlot <http://www.sigmaplot.com/> [letzter Zugriff 30.6.2015].

⁵² Homepage von GraphPad Prism: <http://www.graphpad.com/scientific-software/prism/> [letzter Zugriff 30.6.2015].

3.2.2.5 Beratungen und Schulungen im Softwarebereich (8)

„Als Forschende, wünschen wir uns mehr Beratung und Schulungen zum Thema Software sowie mehr Support im Softwarebereich, um die regelmäßig entstehenden Fragen zur Bedienung unserer Arbeitssoftware zu klären.“

Es wurde von Forschenden eine Reihe von Wünschen zu Beratungen, Schulungen, Informationen und personeller Unterstützung im Softwarebereich ausgesprochen. So wurde bspw. Support bei der Installation von Softwareprodukten gewünscht. Konkret haben mehrere Forschende den Beratungs-, -informations und Schulungsbedarf zur Nutzung der Videobearbeitungssoftware von Adobe⁵³ Software für LaTeX, CAD, GIS sowie von Text- und Audiotextbearbeitungssoftware ausgesprochen.

3.2.2.6 Software zur Datenerhebung (12)

„Als Forschende, wünschen wir uns spezielle Softwarelösungen, um die Erhebung von Daten zu erleichtern.“

In zehn User-Stories haben Forschende Bedarf nach Software zur Datenerhebung und zur Versuchsprotokollierung ausgesprochen. Es wurden auch Anforderungen nach Softwarelösungen geäußert, um Daten nicht auf Papier, sondern direkt auf Rechnern sowie mobilen Geräten (Tablets und Smartphones) eintragen zu können. Um die Datenerhebung bei der Nutzung von Messgeräten effizienter zu gestalten, wurde der Wunsch nach einer Verbesserung der Kommunikationsmöglichkeiten zwischen Messgeräten und den Laborrechnern geäußert.

3.2.2.7 Software zur Datenpublikation (6)

„Als Forschende, wünschen wir uns spezielle Softwarelösungen, um bestimmte Aspekte unserer Daten offen darstellen zu können.“

Nicht für jeden Datensatz findet sich ein geeignetes Repositorium. Hier wurden User-Stories gesammelt, die bspw. eine bestimmte Datenbanktechnologie für Visualisierung benötigen oder darauf hinauslaufen nur Teile von Daten zu publizieren (Code-Fragmente) oder einfach einen Konvertierungsschritt zur Publikation erstreben.

3.2.3 Status quo

An allen neun Universitäten existieren Webportale mit Informationen für Mitarbeiter zum Thema Softwarebeschaffung und Beratung (s. Tab. 1). Auf diesen Webportalen werden Listen von zur Verfügung stehenden Softwareprodukten angeboten. Es werden dabei günstige Landeslizenzen, spezielle Lizenzprogramme, Campuslizenzen, Campuslizenzen für Studierende sowie Open-Source-Lösungen erwähnt. Außer Beratung zur Beschaffung der Software werden Weiterbildungsangebote präsentiert, sowie weiterleitende Informationen und Softwareokumentation aufgelistet. Weitere Informationen zur Softwarebeschaffung sind in 6.3. [Lizenzierung](#) beschrieben.

⁵³ Homepage von Adobe <http://www.adobe.com/de/> [letzter Zugriff 30.6.2015].

Universität	Link zum Softwareportal [letzter Zugriff 30.6.2015]
Freiburg	https://ruf.asknet.de/cgi-bin/home
Heidelberg	http://www.urz.uni-heidelberg.de/service-katalog/software/
Hohenheim	https://kim.uni-hohenheim.de/95430
Karlsruhe	http://www.scc.kit.edu/produkte/index.php
Konstanz	https://www.rz.uni-konstanz.de/angebote/software/
Mannheim	http://www.uni-mannheim.de/rum/ueber_uns/arbeitsgruppen/csi/beschaffung/soft/
Stuttgart	http://www.hlrs.de/organization/sos/
Tübingen	http://www.tik.uni-stuttgart.de/dienste/Lizenzen/
Ulm	http://www.uni-ulm.de/einrichtungen/kiz/it/software-betriebssysteme.html

Tab. 1: Softwareportale an den Universitäten in Baden-Württemberg

3.2.4 Gegenwärtige Handlungsempfehlungen

Entwicklungsprojekte für Spezialsoftware Infrastruktureinrichtungen können die Community nicht flächendeckend mit speziell angefertigter Software versorgen. Hier sollte an jeder Universität ein eigenes Abwägungskonzept aus Breitenwirkung und Spezialanpassung umgesetzt werden, sodass einzelne Spezialprojekte dennoch auch eine Synergiewirkung erzielen. Eine starke Modularisierung der Softwareentwicklung zuzüglich Quelloffenheit, Interoperabilität und Kompatibilität zu Standardformaten sollte von jedem geförderten Entwicklungsprojekt verlangt werden. Dabei könnte man auch ein konkretes allgemeines Format vorgeben das unterstützt werden soll, sodass sich „Standardformate“ herauskristalisieren (beispielsweise für array-, text- oder netzartige Daten), die häufiger auch von „Insellösungen“ unterstützt werden. Die Hoffnung wäre, dass sich durch diese Brücken die vielen Software-„Inseln“ zu verbinden beginnen und durch die Modularität mit wenig Aufwand passgenaue Werkzeuge geschaffen werden können.

Benutzerfreundlichkeit von Spezialsoftware Benutzerfreundlichkeit ist enorm wichtig. Man könnte diese häufig dadurch validieren, dass Software am Ende aktiv einer ähnlichen Community außerhalb von Baden-Württemberg vorgestellt wird und diese dann um einen quantifizierbaren ersten und einen späteren langfristigen Eindruck bittet.

Beratung und Schulungen zum Thema Software

Forschende können bei der Auswahl und Beschaffung von Softwaretools für ihre Forschungsarbeit beratend unterstützt werden. Außerdem wäre eine Erweiterung des Angebotes zu den Softwareschulungen sinnvoll. Dabei könnte Lizenzierung von Videotutorialportalen, wie es an der Universität Hohenheim⁵⁴ organisiert wurde, einen nicht unerheblichen Teil des Schulungs-

⁵⁴ www.video2brain.com und www.lynda.com [letzter Zugriff 30.6.2015].

bedarfs decken. Auch Beratung im Bereich Open-Source-Software könnte Forschenden helfen, die notwendigen Softwarelösungen zu finden und ohne anfallende Kosten zu beschaffen.

3.3 Formate und einheitliche Standards (186)

3.3.1 Einleitung

Einheitliche Formate und Standards ermöglichen, dass unterschiedliche Methoden, Systeme und Dateien miteinander kompatibel sind und flexibel eingesetzt werden können. Standards sind in diesem Sinne bestimmte einheitliche Richtlinien für die Entwicklung und Nutzung von industriellen Erzeugnissen, Methoden, Formaten usw. Einheitliche Formate und Standards bilden eine wichtige Grundlage für ein effizientes Arbeiten. Die Standardisierung von Datenformaten und Schnittstellen wird als eine wichtige Voraussetzung für die Entwicklung von virtuellen Forschungsumgebungen betrachtet.⁵⁵ Einheitliche Formate und Standards spielen auch eine wichtige Rolle für verschiedene Aspekte des Forschungsdatenmanagements⁵⁶.

User-Stories aus den Natur- und Ingenieurwissenschaften sind in diesem Kapitel im Verhältnis zu ihrer Gesamtzahl erheblich überrepräsentiert (vgl. [Abb. 19](#) mit [Abb. 1](#))

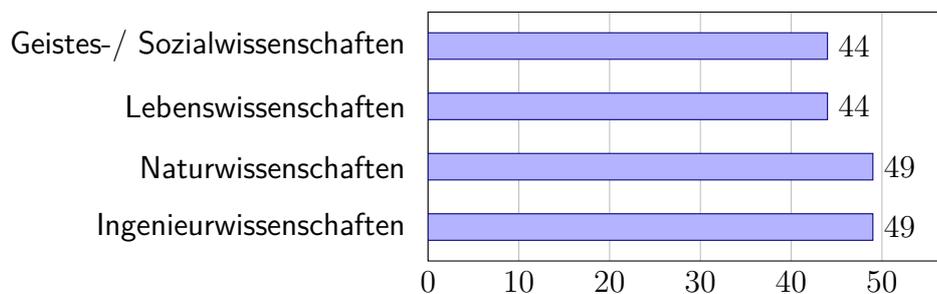


Abb. 19: Verteilung der User-Stories nach der Fachsystematik der DFG

3.3.2 Bedarf der Wissenschaftsgemeinschaft

Um die Bedarfe zur Standardisierung besser zu verstehen, wurde die Frage 3 des Interview-Leitfadens „Mit welchen Datenarten arbeiten Sie hauptsächlich?“ analysiert.

Wie aus [Abb. 20](#) ersichtlich wird, ist der Nutzungsanteil für „fachspezifisches Format (z. B. Messdaten)“ am höchsten. Die Hälfte aller Befragten nutzt in ihrer Arbeit ein solches Datenformat. „Bild, Tabellen, Text“ nutzen etwa 41 % der Befragten. „Fremde Datensätze“ sowie „Datenbanksysteme“ spielen bei wissenschaftlichen Arbeiten mit jeweils ca. 31 und 26 % ebenso eine erhebliche Rolle.

⁵⁵ ESFK, 100–102.

⁵⁶ Büttner, S., Hobohm, H.-C., & Müller, L. (eds.) (2011). Handbuch Forschungsdatenmanagement, Bad Honnef: Bock + Herchen.

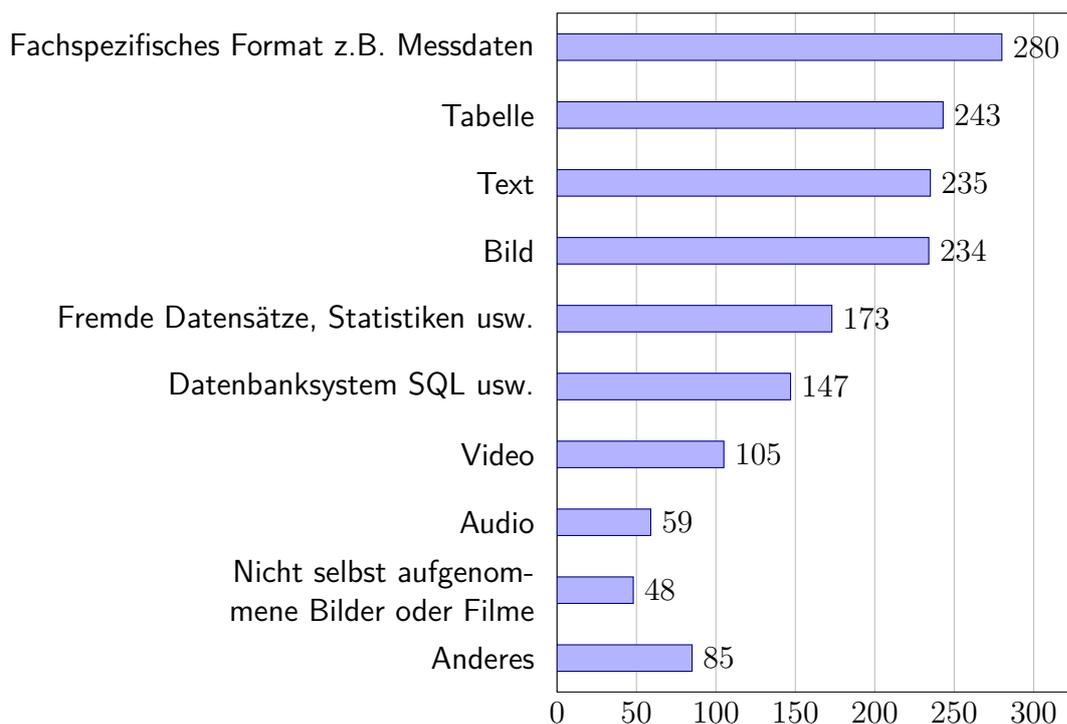


Abb. 20: „Mit welchen Datenarten arbeiten Sie hauptsächlich?“ (Frage 3)

3.3.2.1 Einheitliche Standards (122)

„Als Forschende, welche in eigener Forschungsarbeit mit unterschiedlichen Dateiformaten arbeiten müssen, wünschen wir uns die Einführung bzw. Weiterentwicklung von einheitlichen Standards, um unsere Arbeit mit verschiedenen Datenformaten einfacher und effizienter zu gestalten.“

In diesen User-Stories wurde der Wunsch nach einheitlichen Standards ausgesprochen. Tab. 2 beinhaltet die Auflistung von Zwecken, die einheitliche Standards ermöglichen sollen. Diese Auflistung basiert auf der Auswertung von User-Stories, die einen Bezug zu einheitlichen Standards beinhalten.

Wunsch	Anzahl
Erhöhung der Datenqualität, der Vergleichbarkeit und Verständlichkeit der Daten	28
Vereinfachung des Datenaustauschs	21
Erhöhung der Nachhaltigkeit, sowie Erleichterung der Datennachnutzung	16
Senkung des Aufwands für die Konvertierung, sowie Aufbereitung von Daten	15
Erleichterung der Interoperabilität der Formate	12
Vereinfachung des Publikationsvorgangs	6

Tab. 2: Verteilung der angesprochenen Zwecke, die durch einheitliche Standards erreicht werden sollen. Mehrfachnennungen von User-Stories in mehreren Kategorien sind möglich.

Der Hauptzweck gewünschter Standardisierung ist die Erhöhung der Datenqualität, der Vergleichbarkeit und Verständlichkeit der Daten sowie für die Vereinfachung der Datennachnutzung (28). Ebenso wurde eine Vereinfachung des Datenaustauschs zwischen den Forschenden genannt (19). Auch die Senkung des Aufwands für die Konvertierung sowie für die Aufbereitung von Daten Anderer für eine Weiternutzung in eigenen Studien wurde als Grund für den Wunsch nach Standardisierung erwähnt (17). Interoperabilität der Formate, also die Möglichkeit gleiche Formate in unterschiedlichen Anwendungen einzusetzen, wurde in 12 User-Storys als Beweggrund für eine Standardisierung angesprochen. Ebenso wurde der Wunsch nach Nachhaltigkeit sowie einer Nachnutzung der Formate erwähnt (16). Weiterhin wurde das Problem angesprochen, dass manche wertvolle, vor Jahren erstellte Daten nicht mehr nutzbar sind, da keine Software mehr existiert, welche diese lesen und bearbeiten kann.

In 6 Fällen wurde ein Bedarf nach Standardisierung der Forschungsdaten sowie der Manuskriptformate geäußert, um den Veröffentlichungsaufwand zu senken.

Betrachtet man näher was genau standardisiert werden soll, so wurde in 53 User-Storys der Wunsch nach einheitlichen Metadatenstandards für Datenbeschreibung erwähnt. Davon erhoffen sich 40 Befragte eine Erhöhung des Werts der Daten. In einem Satz zusammengefasst zeigen sich folgende Zwecke, für die Metadaten als nützlich angesehen werden:

„Als Forschende, wünschen wir uns einheitliche Metadatenstandards für die Datenbeschreibung, um einfachere und effizientere Nutzung von Forschungsdaten, inkl. Datenauffindbarkeit, -zugriff, -verständlichkeit, -vergleichbarkeit, -aufbereitung, -konvertierung, -verwaltung, -austausch, -analyse, -interpretation, -reproduzierbarkeit, und -nachnutzbarkeit zu ermöglichen.“

Es wurde auch ein Bedarf nach standardisierten Schnittstellen genannt, damit der Zugriff auf Daten aus unterschiedlichen Datenquellen, Softwaresystemen, Geräten usw. vereinfacht wird (11). Wünsche zur Datenkonvertierung wurden in 17 User-Storys erwähnt. Dabei wurden in neun Fällen Voraussetzungen gewünscht, um den Konvertierungsaufwand zu reduzieren bzw. eine Konvertierungsnotwendigkeit ganz zu vermeiden. In neun Fällen wurde ein Bedarf nach Unterstützung bzw. nach speziellen Tools ausgesprochen, um den Konvertierungsvorgang zu ermöglichen und effizienter zu gestalten (vgl. Tab. 3).

Wunsch	Anzahl
Summe der Wünsche zur Datenkonvertierung	17
Voraussetzungen, um den Konvertierungsaufwand zu reduzieren bzw. Konvertierungsnotwendigkeit ganz zu vermeiden	9
Unterstützung, bzw. spezielle Tools, um den Konvertierungsvorgang zu ermöglichen bzw. effizienter zu gestalten	9
Standardisierte Schnittstellen, um Zugriff auf Daten aus unterschiedlichen Datenquellen, Softwaresystemen, Geräten usw. zu vereinfachen	11

Tab. 3: Verteilung der angesprochenen Bedarfe sowie deren Zwecke. Mehrfachnennungen von User-Storys in mehreren Kategorien sind möglich.

3.3.2.2 Softwarewünsche (85)

„Als Forschende, welche in der eigenen Forschung auf die Nutzung von bestimmten Softwareprodukten und den dazugehörigen Datenformaten angewiesen sind, wünschen wir uns mehr kompatible Softwarelösungen, um ohne problematische Konvertierungen forschen zu können.“

Die User-Stories dieses Abschnitts überschneiden sich stark (50) mit [3.2. Wissenschaftliche Softwaretools](#). Die Schnittmenge ist der Ähnlichkeit der Anforderungen bzw. schweren Trennbarkeit von Anforderungen an Software von denen an ihre Schnittstellen geschuldet. Am häufigsten wurden Bedarfe nach Softwarelösungen zur Datenaufbereitung, -verwaltung und Metadatenenerstellung erwähnt. Sie umfassten bspw. die Einführung von Laborjournal-Software, um experimentelle Abläufe elektronisch zu erfassen und zu verwalten. Zusätzlich wurden Softwaretools zur Konvertierung von unterschiedlichen Formaten untereinander gewünscht (23).

Auch der Wunsch nach mehr Einsatz von Open-Source-Softwarelösungen und Formaten wurde geäußert (10). Offene Formate ermöglichen den Einsatz von verschiedenen frei zugänglichen sowie selbst entwickelten Softwarelösungen zur Nutzung dieser Dateiformate. Das erweitert ihre Einsatzmöglichkeiten und führt zur Unabhängigkeit von kostenpflichtigen Softwareprodukten, die im Fall der Nutzung von proprietären Formaten gegen Entgelt beschafft werden müssen.

Es wurde 14 mal explizit Bedarf zur Verbesserung der Kompatibilität zwischen unterschiedlichen Softwarelösungen geäußert. So sollten Softwareprodukte, die mit ähnlichen bzw. gleichen Dateiformaten arbeiten, weniger Kompatibilitätsprobleme untereinander aufweisen. Als eine mögliche Lösung zur Verbesserung der Kompatibilität bei Nutzung von Softwareprodukten wurde die Option vorgeschlagen, Forschenden an einer Universität gleiche Softwareprodukte und gleiche Versionen der Arbeitssoftware zur Verfügung zu stellen, um so Inkompatibilitäten zu vermeiden.

Es gab auch Wünsche zur Verbesserung der Benutzerfreundlichkeit (7) und Nachhaltigkeit (16) von Softwareprodukten. Unter anderem wurde gewünscht, dass bei einer neuen Softwareversion die alte Programmoberfläche beibehalten wird, um den Aufwand des Erlernens einer neuen Softwareoberfläche zu umgehen.

3.3.2.3 Wünsche nach Richtlinien (38)

„Als Forschende wünschen wir uns klare Richtlinien, um die zur Verfügung stehenden Datenformate und Standards effizient und im Einklang mit aktuellen Anforderungen der wissenschaftlichen Community einsetzen zu können.“

Um vorhandene Formate und Standards effizienter in der eigenen Arbeit nutzen zu können, sind entsprechend den Aussagen von Forschenden klare und einheitliche Richtlinien notwendig (s. [Tab. 4](#)). Als Institutionen, die diese Richtlinien entwickeln sollten, wurden bspw. das MWK und die DFG genannt.

Insgesamt wurden Richtlinien zum Umgang mit Forschungsdaten in 26 User-Stories ausgesprochen. Unter diesen User-Stories gab es Wünsche nach Richtlinien bzw. Leitfäden

Wunsch	Anzahl
Summe der Wünsche nach klaren Richtlinien	26
Richtlinien zur Datenaufbereitung	15
Richtlinien zur Datenspeicherung, inkl. Sicherung und Archivierung	15
Richtlinien zur Datenbeschreibung und Metadatenerstellung	13
Richtlinien zur Erhöhung der Datenqualität, inkl. Verständlichkeit und Nachnutzbarkeit, Einheitlichkeit, Vergleichbarkeit, Auffindbarkeit usw.	10
Richtlinien zur Datenpublikation	6
Richtlinien zur Einhaltung von bestimmten Standards, um Datenaustausch zu vereinfachen	3
Richtlinien zur Softwarebeschreibung, um ihre Nachnutzbarkeit zu gewährleisten	1

Tab. 4: Verteilung der angesprochenen Wünsche nach Richtlinien. Mehrfachnennungen von User-Stories in mehreren Kategorien sind möglich.

zur Datenaufbereitung (15), zur Datenspeicherung inkl. Backup und Archivierung (15), zur Datenbeschreibung und Metadatenerstellung (13) sowie Richtlinien zur Datenpublikation (6).

Die Richtlinien zur Einhaltung bestimmter Standards sollten dabei den Austausch von Daten zwischen verschiedenen Forschungsgruppen vereinfachen (3) sowie zur Erhöhung der Datenqualität, inklusive Verständlichkeit, Nachnutzbarkeit, Einheitlichkeit, Vergleichbarkeit und Auffindbarkeit, beitragen (10).

Gewünscht wurden nicht nur klare, sondern auch verpflichtende Richtlinien. Dabei wurde auch ein Bedarf nach entsprechenden finanziellen und zeitlichen Ressourcen erwähnt, um diese Richtlinien in vollem Maße befolgen zu können.

3.3.3 Status quo

Die rapide und breit verstreute Entwicklung im Bereich Informatik erfordert die Einführung von Standards⁵⁷. Dazu gibt es eine Reihe von Gremien, die an der Entwicklung von Empfehlungen, Standards und Normungen beteiligt sind^{58,59}. Diese Normungsgremien sind in regionalen, nationalen und internationalen Organisationen vertreten.

Auf der internationalen Ebene spielt die ISO - *International Organization for Standardization*⁶⁰ eine besonders wichtige Rolle. Die ISO ist ein internationales Gremium, welches von der UNESCO gegründet wurde. Von der ISO werden Empfehlungen ausgesprochen sowie Normen definiert. Die IEC⁶¹ - *International Electrotechnical Commission* ist eine Non-Profit-Organisation mit einer sehr langen Geschichte, die an der Schnittstelle zwischen Forschung und Industrie zur Entwicklung von zahlreichen Standards auf dem Gebiet der Elektrotech-

⁵⁷ Taschenbuch der Informatik, Uwe Schneider (Hrsg.), Carl Hanser Verlag München, 2012.

⁵⁸ Computerschnittstellen und Bussysteme, Dembowski, K., 3. völlig neu bearbeitete Auflage, VDE VERLAG GMBH 2013.

⁵⁹ Taschenbuch der Informatik, Uwe Schneider (Hrsg.), Carl Hanser Verlag München, 2012.

⁶⁰ ISO - *International Organization for Standardization* <http://www.iso.org> [letzter Zugriff 28.6.2015].

⁶¹ IEC - *International Electrotechnical Commission* <http://www.iec.ch> [letzter Zugriff 28.6.2015].

nik, Elektronik und Informationstechnologie beigetragen hat. Das JTC1⁶² - *Joint Technical Committee 1* repräsentiert ein vereintes Gremium von ISO und IEC zur Koordinierung ihrer Standardisierungstätigkeiten im Bereich der Informationstechnologie.

Zu den wichtigen Europäischen Standardisierungsorganisationen gehört das CEN - *Comité Européen de Normalisation*⁶³. Das CEN repräsentiert einen Zusammenschluss von europäischen nationalen Standardorganisationen. Das ETSI - *European Telecommunications Standards Institute*⁶⁴ entwickelt als Vereinigung von europäischen Verwaltungen, Forschungseinrichtungen und Betrieben Standards für Informationstechnologie und Telekommunikation.

Manche nationale Normungsgremien wie bspw. das ANSI - *American National Standards Institute*⁶⁵ sind auch für die weltweite Standardisierung von hoher Relevanz. Zu ANSI gehören eine Reihe von amerikanischen Organisationen, die Standards für den international weit verbreiteten ANSI-Zeichensatz festlegen. Außerdem spricht diese Organisation Empfehlungen aus, die in vielen Fällen als IEEE-Standards festgelegt werden. Das IEEE - *Institute of Electrical and Electronic Engineers*⁶⁶ ist ein Zusammenschluss von amerikanischen Firmen, der weltweit relevante Standards festlegt.

In Deutschland gehört das DIN - *Deutsches Institut für Normung*⁶⁷ zu den wichtigsten Standardisierungsorganisationen. Eine nationale Standardisierungsorganisation ist die DKE - *Deutsche Kommission Elektrotechnik Elektronik Informationstechnik*⁶⁸. Die DKE entwickelt, ebenso wie der NIA - *Normenausschuss Informationstechnik und Anwendungen*⁶⁹, Normen im Bereich der Informationstechnik. Als Informationsquellen zum Thema Normen und Standards stehen in Deutschland Veröffentlichungen des Beuth Verlags⁷⁰ sowie die Webseite des Deutschen Instituts für Normung⁷¹ zur Verfügung.

Proprietäre Standards werden oft von Herstellerfirmen durch den Verkauf ihrer kommerziellen Lösungen verbreitet. Es gibt jedoch auch vermehrte Bestrebungen zur Entwicklung von offener Standards. Solche Standards können frei oder unter den Bedingungen einer moderaten, nicht diskriminierenden Lizenzierung zur Verfügung gestellt werden. Einige große Standardisierungsorganisationen beteiligen sich an der Entwicklung offener Standards. Dazu gehören u. a. das bereits erwähnte IEEE⁷², das W3C - *WorldWideWeb Consortium*⁷³, die IETF - *Internet Engineering Task Force*⁷⁴ und das IAB - *Internet Architecture Board*⁷⁵.

⁶² JTC1 - *Joint Technical Committee 1* <http://www.jtc1.org> [letzter Zugriff 28.6.2015].

⁶³ CEN - *Comité Européen de Normalisation* <http://www.cenorm.be> - [letzter Zugriff 28.6.2015].

⁶⁴ ETSI - *European Telecommunications Standards Institute* <http://www.etsi.org> [letzter Zugriff 28.6.2015].

⁶⁵ ANSI - *American National Standards Institute* <http://www.ansi.org> [letzter Zugriff 28.06.2015]

⁶⁶ IEEE - *Institute of Electrical and Electronic Engineers* <http://www.ieee.org> [letzter Zugriff 28.6.2015].

⁶⁷ DIN - *Deutsches Institut für Normung* <http://www.normung.din.de> - [letzter Zugriff 28.6.2015].

⁶⁸ DKE - *Deutsche Kommission Elektrotechnik Elektronik Informationstechnik* <http://www.en.din.de> - [letzter Zugriff 28.6.2015].

⁶⁹ NIA - *Normenausschuss Informationstechnik und Anwendungen* <http://www.ni.din.de> - [letzter Zugriff 28.6.2015].

⁷⁰ Beuth Verlag - Aktuelle Normen und Fachliteratur <http://www.beuth.de> - [letzter Zugriff 28.6.2015.]

⁷¹ Webseite des Deutschen Instituts für Normung <http://www.din.de> - [letzter Zugriff 28.6.2015].

⁷² IEEE - *Institute of Electrical and Electronic Engineers* <http://www.ieee.org>

⁷³ W3C - *WorldWideWeb Consortium* <http://www.w3c.org> - [letzter Zugriff 28.6.2015].

⁷⁴ IETF - *Internet Engineering Task Force* <http://www.ietf.org> - [letzter Zugriff 28.6.2015].

⁷⁵ IAB - *Internet Architecture Board* <http://www.iab.org> - [letzter Zugriff 28.6.2015].

3.3.4 Gegenwärtige Handlungsempfehlungen

Weiterentwicklung der Konzepte zur Optimierung und Vereinheitlichung von Daten- und Softwarestandards.

Aufgrund der Uneinheitlichkeit der angewendeten Formate sowie des Fehlens einheitlicher Standards, können während der Forschungsarbeit erhebliche Zeitverluste und Dateninformationsverluste auftreten. Um diese Verluste zu reduzieren, sind folgende Lösungsansätze vorstellbar:

- Förderung der Nutzung von einheitlichen Softwareprodukten durch Forschende
 - durch Beschaffung und Zurverfügungstellung einheitlicher Softwarelizenzen (Campus- oder Landeslizenzen).
 - mehr Unterstützung von Open-Source-Formaten und Programmen an den baden-württembergischen Landesuniversitäten durch Beratung- und Supporterweiterung. Open-Source-Software ist häufig sehr leistungsfähig und die Universitäten sollten abwägen, ob Lizenzierungskosten tatsächlich immer niedriger sind als in Beratung und Entwicklungshilfe von Open-Source-Alternativen zu investieren.
- Förderung der Nachhaltigkeit für Softwareprodukte und Datenformate durch einen längeren Support von gleichen Softwareprodukten sowie deren Versionen, zumindest solange keine Bedenken bezüglich fehlender Sicherheitsupdates existieren.
- Stärker bevorzugte Förderung bei Verwendung offener Schnittstellen und reproduzierbarer Funktionalität von Software und Geräten.

Erweiterung der Beratungs- und Informationsangebote zum Thema „Standards, Formate und deren Konvertierung“ durch:

- Einrichtung einer Beratungsstelle und einer Webplattform zum Umgang mit Standards und Formaten.
- Erarbeitung von Richtlinien zum Umgang mit Standards und Formaten sowie mit Forschungsdaten insgesamt.

4 Datenerhebung und -austausch

4.1 Daten außeruniversitärer Einrichtungen (117)

4.1.1 Einleitung

In vielen User-Stories wurde der Zugang zu außeruniversitären Daten problematisiert. Dabei handelt es sich um eine Vielzahl sehr heterogener Daten von unabhängigen oder wissenschaftsfernen Institutionen, auf die die Forschenden gern zurückgreifen möchten. Dass diese Daten nicht immer für die Forschenden zur Verfügung stehen, ist hierbei nur ein prominentes der vielen Probleme dabei. Auch die Weiterverarbeitung von gefundenen und prinzipiell nutzbaren Daten, bereitet den Forschenden oft einige Mühe.

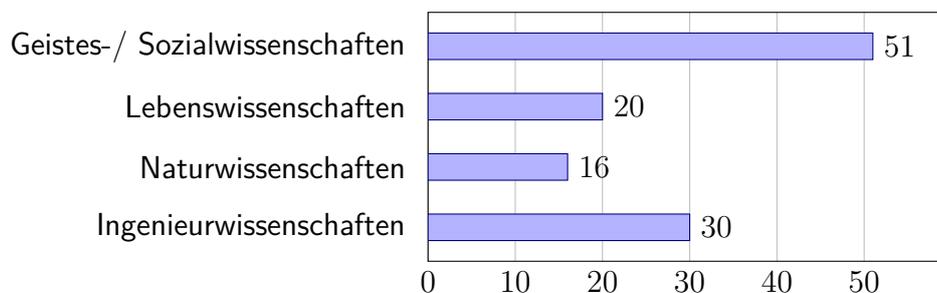


Abb. 21: Verteilung der User-Stories nach der Fachsystematik der DFG

Mit 51 User-Stories kommen die meisten der 117 User-Stories aus den Geistes- und Sozialwissenschaften. Dieser Umstand überrascht kaum, weil in vielen Disziplinen der Geistes- und Sozialwissenschaften die Forschung auf „externen Daten“ beruht, und in den anderen Fachbereichen weitaus häufiger die Forschungsdaten selbst oder von anderen Forschenden erhoben werden. Auffällig ist die, gegenüber ihrem Vorkommen in der Befragungsstichprobe, deutlich überproportionale Häufung der Ingenieurwissenschaften (vgl. Abb. 21 mit Abb. 1). Diese treten hier oft in ihrer direkten Relevanz für die Gesellschaft in Erscheinung, vor allem mit Bezug zum Bauwesen und Architektur (Stadtplanung, Verkehrswesen usw.) aber auch im Kommunikations- und Elektrotechnikbereich.

4.1.2 Bedarf der Wissenschaftsgemeinschaft

Die 117 User-Stories zeichnen ein sehr heterogenes Bild, in dem viele verschiedene Aspekte der Nachnutzungsthematik zum Vorschein kommen. Neben dem allgemeinen Wunsch nach mehr Rohdaten bzw. forschungsrelevanten Daten wird in den User-Stories auch die Notwendigkeit einer Standardisierung zur Erzielung von Nachvollziehbarkeit und die Auffindbarkeit

der forschungsrelevanten Daten angesprochen. Die Schnittmenge mit den anderen Kategorien ist sehr hoch. Bei 58 User-Stories kommt es zu Überschneidungen mit anderen Themen. Die mit Abstand größten Schnittmenge gibt es mit [5.2. Repositorien für die Publikation von Forschungsdaten](#) (17) und [3.3. Formate und einheitliche Standards](#) (13). Es folgt [2.1. Urheberrecht und Datenschutz](#) (7). Diese recht auffälligen Überschneidungen spiegeln sich auch ein wenig in der Kategorisierung der User-Stories in Unterthemen wider.

4.1.2.1 Offene Verwaltungsdaten (13)

„Als Forschende, wünschen wir uns, dass öffentliche Institutionen im Sinne der Open-Government-Data-Initiative ihre erhobenen Daten für die Wissenschaft kostenlos und unbürokratisch in geeigneten Datenformaten zur Verfügung stellen, um diese Daten für unsere Forschung effektiv nutzen zu können.“

In sieben User-Stories wünschen sich die Forschenden ganz allgemein, dass öffentliche Institutionen, die mit Steuergeldern finanziert werden, ihre erhobenen Daten unbürokratisch und kostenlos bzw. zu geringen Kosten öffentlich und leicht zugänglich anbieten. Aus den User-Stories lässt sich herauslesen, dass es sich in erster Linie um Daten der Behörden in den Gemeinden und Kommunen handelt. In einer weiteren User-Story wird ein statistische Landesamt explizit erwähnt. Dazu kommen drei weitere User-Stories, in denen der Wunsch geäußert wird auf geographische Daten zugreifen zu können. Zweimal wird in den User-Stories das Landesamt für Geoinformation und Landentwicklung explizit genannt. In den übrigen User-Stories wird der Wunsch nach öffentlich zugänglichen Daten im Zusammenhang mit einzelnen Ämtern, Behörden und anderen Institutionen genannt. Außerdem wird in einigen User-Stories ebenso deutlich, dass die gewünschten Daten in maschinenlesbaren Formaten vorliegen sollen, um sie für die Forschung nicht mühevoll aufarbeiten zu müssen. Dieser Wunsch kam ebenfalls bei einigen User-Stories in [4.1.2.3. Standardisierung von Forschungsdaten von Software](#) vor.

4.1.2.2 Einfacher Zugang zu nicht-offenen Daten (64)

„Als Forschende, wünschen wir uns einen unbürokratischen sowie finanziell und rechtlich klar geregelten Zugang zu potenziellen Forschungsdaten anderer Institutionen und der freien Wirtschaft, um die Daten ohne große Verzögerungen für unsere Forschungsvorhaben nutzen zu können.“

Die meisten User-Stories zum Zugang zu außeruniversitären Daten können im Wunsch nach einem einfachen bzw. klar geregelten und unbürokratischen Zugang zu forschungsrelevanten Daten zusammengefasst werden und lassen sich dabei in zwei Gruppen unterteilen.

In 36 User-Stories wünschen sich die Forschenden Zugang zu Daten von öffentlichen Institutionen, die nicht eindeutig der Verwaltung zugeordnet werden können. Im Unterschied zu den User-Stories aus [4.1.2.1. Offene Verwaltungsdaten](#) gibt es bei diesen forschungsrelevanten Daten Beschränkungen (beispielsweise Sperrfristen auf Dokumente, Fernsehbeiträge, sicherheitsrelevante Daten von Reaktoren), die eine öffentliche Bereitstellung unmöglich machen. Dies ist den Forschenden durchaus bewusst, sodass hier in erster Linie eine Vereinfachung des Antragsweges zur Einsicht der Daten gewünscht wird. Dass diese Daten für die Wissenschaft schließlich kostenlos zur Verfügung gestellt werden, ist ein weiterer Aspekt, der sich

in mehreren User-Stories wiederfindet. In den User-Stories kommen vereinzelt auch sehr klar formulierte Wünsche vor. Aus einem kleinen Fachgebiet kam der Wunsch, dass die leichtere Zugänglichkeit zu den Magister- und Masterarbeiten des Faches, die oft die erste und einzige Auseinandersetzung mit einem Thema darstellen, die Forschenden zeitlich und bürokratisch entlastet. Diesem und ähnlichen Wünschen sollte universitätsintern nachgegangen werden.

In den 28 anderen User-Stories dieser Unterkategorie wird die Zugänglichkeit zu Daten aus der Wirtschaft angesprochen. Auch hier wünschen sich die Forschenden einen einfachen und finanziell klar geregelten Zugang zu den notwendigen Daten. In einigen User-Stories kommt zu Tage, dass die Forschenden sich hier eine bürokratische Vereinfachung erhoffen, um Geld aus ihren Haushaltsmitteln für den Ankauf verwenden bzw. um Geldmittel für den Ankauf beantragen zu können.

4.1.2.3 Standardisierung von Forschungsdaten von Software (21)

„Als Forschende, wünschen wir uns eine verstärkte Standardisierung der offen zugänglichen Forschungsdaten im Sinne von Dateiformaten, maschineller Weiterverarbeitung und Schnittstellen an Geräten, um die Forschungsdaten ohne große Aufbereitung für unsere eigene Forschung nachnutzen zu können.“

Die konsequente Verfolgung von Standards von Dateiformaten und Software im Zusammenhang mit Forschungsdaten wird bei 21 User-Stories deutlich. In sechs User-Stories wird der Wunsch nach der Vereinheitlichung und Reduzierung von Dateiformaten deutlich, wenn öffentliche Institutionen forschungsrelevante Daten anbieten. Darüber hinaus wird in drei weiteren User-Stories auf maschinenlesbare und offene Dateiformate Wert gelegt, um die Daten ohne zusätzlichen Aufwand weiterverarbeiten zu können. Der Wunsch nach nicht-kommerziellen und offenen Systemen zeigt sich auch bei drei anderen User-Stories, in denen sich (langlebige) Open-Source-Software bzw. standardisierte Schnittstellen an Messinstrumenten gewünscht werden, um nicht auf bestimmte kommerzielle Produkte und Anbieter angewiesen zu sein.

4.1.2.4 Auffinden von forschungsrelevanten Daten (19)

„Als Forschende, wünschen wir uns eine erhöhte Transparenz der, im Internet an vielen verschiedenen Orten bereitgestellten, forschungsrelevanten Daten durch betreute Portale und Metadatenbanken, um relevante Daten leichter finden zu können.“

Dass das Auffinden von prinzipiell zugänglichen Forschungsdaten oft zeitintensiv ist, kommt in vielen User-Stories dieses Kapitels vor. Die Forschenden erhoffen sich durch Portale und Ähnliches (Metadatenbanken, Broschüren, usw.) einen einfacheren und bequemerem Überblick über bestimmte Forschungsgebiete zu erhalten. Allein 14 User-Stories geben einen solchen Wunsch zentral wieder. Hinzu kommen vier User-Stories, in denen eine Standardisierung der Metadaten die Suche vereinfachen soll. Gewissermaßen einen Spezialfall zeigt der Wunsch einer User-Story, die persönlichen Daten potentieller (medizinischer) Probanden zentral zu verwalten, um zügig spezifisch passende Testgruppen zusammenzustellen.

4.1.3 Status quo

Daten aus der öffentlichen Verwaltung – Open Government Im Strategiepapier des Bundes zur Open-Government-Initiative wird die Wissenschaft eindeutig als Zielgruppe genannt und die verbundenen Synergieeffekte durch die Bereitstellung öffentlicher Daten angeführt. Eine Priorität bei der Zugänglichmachung erhalten dabei Umweltdaten, geographische Daten, Verkehrsdaten, Daten aus Verwaltung und Politik sowie statistische Daten.⁷⁶ Wenngleich in jenem Bericht auch Empfehlungen zur Standardisierung von Formaten und Metadaten ausgesprochen werden, ist die Erfassung des Bestandes öffentlicher Daten und Formate noch lange nicht abgeschlossen und in den jeweiligen Bundesländern daher auch sehr unterschiedlich.⁷⁷ Bundesbezogene Daten können zum Teil in maschinenlesbarer Form über das Datenportal *GovData*⁷⁸ abgerufen werden. Diese Webseite informiert auch über Metadaten, Dateiformate und Lizenzmodelle. Die bereits enthaltenen Daten sind zumeist in verschiedenen Dateiformaten verfügbar, meist auch in benutzerfreundlichen Formaten wie CSV oder XML. Außerdem werden Richtlinien und Empfehlungen für Nutzer, aber auch für „Bereitsteller“ von Daten zur Verfügung gestellt.

Daten aus der Wissenschaft Viele wissenschaftliche Datensammlungen können über das Portal *Registry of Research Data Repositories*⁷⁹ gefunden werden. Hier findet sich eine Sammlung einer Vielzahl von Daten unterschiedlicher Fachbereiche aus verschiedenen Ländern. Das Portal leistet einen wichtigen Beitrag für eine strukturierte Auffindbarkeit offener Daten.

Daten, die im Rahmen wissenschaftlicher Projekte entstehen, sind meist nicht öffentlich abrufbar. Manche Arbeitsgruppen ergreifen aber Eigeninitiative und stellen ihre Auswertungsparameter bzw. Publikationen über eigene Webseiten zur Verfügung.⁸⁰ Solche Datensätze sind zwar grundsätzlich vorhanden, jedoch nur schwer auffindbar, da eine strukturierte Abfrage und Suchmöglichkeit nach ihnen fehlt.

⁷⁶ Open Government Data Deutschland. Eine Studie zu Open Government in Deutschland im Auftrag des Bundesministerium des Innern (Juli 2012), http://www.bmi.bund.de/SharedDocs/Downloads/DE/Themen/OED_Verwaltung/ModerneVerwaltung/opengovernment.pdf?__blob=publicationFile, S. 72-74 und S. 390-392 [letzter Zugriff 11.5.2015].

⁷⁷ Für Baden-Württemberg: http://www.verwaltungsreform-bw.de/PUBLIKATIONEN/Studien-Konzepte/Documents/131216_E-GK-Standards%202013.pdf [letzter Zugriff 11.5.2015]. Als Best-Practice-Beispiele nennt der Bericht des BMI (s. Fußnote oben [1], S. 418) u. a. das Open-Data-Portal der Stadt Berlin (daten.berlin.de) und das Portal für eine Geodateninfrastruktur in Deutschland (<http://www.geoportal.de/DE/Geoportal/geoportal.html?lang=de> [letzter Zugriff 11.5.2015]). Als internationale Beispiele werden das europäische Datenportal (publicdata.eu) und das Datenportal der Stadt Wien (<https://www.data.gv.at/> [letzter Zugriff 11.5.2015]) angeführt.

⁷⁸ <https://www.govdata.de/> [letzter Zugriff 11.5.2015].

⁷⁹ <http://www.re3data.org/> [letzter Zugriff 11.5.2015].

⁸⁰ Beispielhaft sei hier auf die Seite von Prof. Grammig, Universität Tübingen, verwiesen: <http://www.wiwi.uni-tuebingen.de/lehrstuehle/statistik-und-oekonometrie/econometrics1/team/prof-dr-joachim-grammig/publications.html> [letzter Zugriff 11.5.2015].

Ähnliche Möglichkeiten bieten im Bereich der Naturwissenschaften auch schon viele Journale und Onlinepublikationen, sodass neben den Ergebnissen in Form von Zusatzmaterialien noch weitere Informationen (Daten, Parameter etc.) zur Verfügung gestellt werden können.⁸¹

Ein weiterer Aspekt, der sich mit den User-Stories in 6.5. Open Science überschneidet, ist natürlich auch, dass die Zugänglichkeit zu Open-Access-Publikationen und die Bereitstellung von Open-Access-Publikationen an allen Standorten in der Verantwortung der Universitätsbibliotheken liegen. Daher soll an dieser Stelle nur kurz auf den Aspekt der sogenannten „Selbstarchivierung“ eingegangen werden. Ein von Forschenden aus naturwissenschaftlichen Fachbereichen besonders häufig genutztes Instrument ist *arXiv.org*⁸², mittels dessen eigene Publikationen aber auch Daten veröffentlicht werden können.

4.1.4 Gegenwärtige Handlungsempfehlungen

Einigen der geäußerten Wünsche kann von Seiten der Universitäten und im Rahmen von Landesprojekten nicht direkt nachgegangen werden, weil in den Wünschen (forschungsrelevante) Daten Dritter angesprochen wurden. Anders liegt der Fall bei Daten, die von Landeseinrichtungen, Behörden und Ämtern, erhoben wurden. Hier gibt es die Möglichkeit, gewissermaßen mit gutem Beispiel voranzugehen und auf die Wünsche der Forschenden einzugehen.

Öffentliche Bereitstellung von Daten – Open Government Im Sinne des Open Government und der Open-Government-Data-Initiative ist eine prinzipielle öffentliche Bereitstellung von Daten aus der Verwaltung begrüßenswert, weil dies zu einer größeren Teilhabe des Einzelnen an gemeinschaftlichen Belangen führen kann. In der Konsequenz sollten die Daten in offenen Formaten bereitgestellt werden. Über die reine Bereitstellung hinaus zeigen die User-Stories, dass die Forschenden besondere Ansprüche an die Form der Daten haben, um diese ohne Zusatzaufwand für ihre Forschung nutzen zu können. Die bereitgestellten Daten sollten also in vereinheitlichten und maschinenlesbaren Formaten vorliegen. Ein geringe Auswahlmöglichkeit an Formaten (strukturierte Daten beispielsweise *.csv, *.ods, *.xml) würden die Forschenden begrüßen.

Auffindbarkeit von Daten – Zentrale Portale Neben dem generellen Zugang zu Daten zeigen die User-Stories auch, dass mitunter bereitgestellte Daten nicht gut gefunden werden. Dem Wunsch nach einem zentralen Webportal, um Daten der Behörden und Ämter abzurufen, sollte entsprochen werden. Über eine solche Seite können die genauen Adressen des Downloadbereichs der einzelnen Einrichtungen verlinkt werden. Diese Vernetzung von öffentlich zugänglichen Daten mit einer ausführlichen Metadatenbeschreibung erhöht die Transparenz dieser Daten und verringert den Rechercheaufwand der Forschenden. Bei gemeinsamer Umsetzung mit der Handlungsempfehlung würde das Land Baden-Württemberg in Deutschland

⁸¹ Zum Beispiel die Richtlinien für Publikationen bei den Verlagen Elsevier: <http://www.elsevier.com/about/content-innovation/inline-supplementary-material-for-journal-articles> [letzter Zugriff 11.5.2015] oder AIP: <http://publishing.aip.org/authors/supporting-data> [letzter Zugriff 11.5.2015].

⁸² <http://arxiv.org/> [letzter Zugriff 11.5.2015].

eine Vorreiterrolle in der Öffnung von Regierung und Verwaltung gegenüber der Bevölkerung einnehmen.

Unbürokratischer und standardisierter Antragsweg Für Daten, die aus verschiedenen Gründen nicht öffentlich bereitgestellt werden können oder sollen, wünschen sich die Forschenden einen Abbau der Bürokratie, sodass die Wartezeit auf forschungsrelevante Daten sichtlich verkürzt wird. Ein vereinheitlichter Antragsweg für die Daten vergleichbarer Einrichtungen löst eine Vielzahl der in den User-Stories geäußerten Missstände. Auch hier bietet sich für das Land Baden-Württemberg die Möglichkeit, auf die Behörden und Ämter Einfluss zu nehmen, um den Forschenden die Beschaffung der Datengrundlage zu vereinfachen.

Zugang per Gesetz Da die getrennten Zuständigkeiten verschiedener Behörden den Einfluss einzelner unterstützender Ministerien leider sehr erschweren, plädieren wir für einen im Gesetz verankerten Rechtsanspruch von Forschenden, um einheitlichen Zugang auf Daten öffentlicher Einrichtungen zu erhalten, soweit dadurch keine Rechte Dritter eingeschränkt werden. Diesen Anspruch kann man aus der Forschungsfreiheit ableiten und alle öffentlichen Behörden und Ämter wären zur Anwendung verpflichtet.

4.2 Erwartungen an die Digitalisierung (41)

4.2.1 Einleitung

Der Begriff „Digitalisierung“ und die Ergebnisse des Prozesses, die Digitalisate, scheinen zunächst als eine Reproduktion von analogen Objekten in digitaler Form eindeutig definiert zu sein. Diese Minimaldefinition birgt jedoch mehrere Lücken, sodass höchst unterschiedliche Ergebnisse, beispielsweise bezüglich Formaten und Maschinenlesbarkeit, bei der Digitalisierung entstehen können, von denen einige Umsetzungen dem wissenschaftlichen Anspruch kaum genügen und für die Forschenden schließlich den einzigen Vorteil haben, „zeit- und ortsunabhängig mit dem kulturellen historischen Erbe zu arbeiten“⁸³. In dieser Bemerkung steckt natürlich ein deutlicher Hinweis, dass die Digitalisierung für historisch arbeitende Fächern bei weitem die größte Relevanz hat. So überrascht es nicht, dass mit 32 User-Stories die große Mehrheit aus geisteswissenschaftlichen Fächern, insbesondere der Geschichtswissenschaft stammt.

Die User-Stories in diesem Kapitel zeigen eindeutig, dass Digitalisate in Bilddateiformaten nicht die einzige gewünschte Form von Digitalisaten sind. Den genauen Wünschen der Forschenden, wie sie in den User-Stories erscheinen, und den Erwartungen und Ansprüchen an die Digitalisierung widmen sich die folgenden Seiten.

⁸³ ESFK, 49. Dass durch Digitalisate unbekanntes oder schwer zugängliches Material der Forschung sichtbar und zugänglich gemacht werden kann, Reise- und Kopierkosten gespart und die Originalbestände geschont werden, sind weitere Vorteile, die im Bericht des MWK angeführt werden. Die Sichtbarmachung und Zugänglichkeit werden aber in erster Linie durch die leichte Auffindbarkeit im Internet und nur bedingt durch die Digitalisierung selbst erreicht.

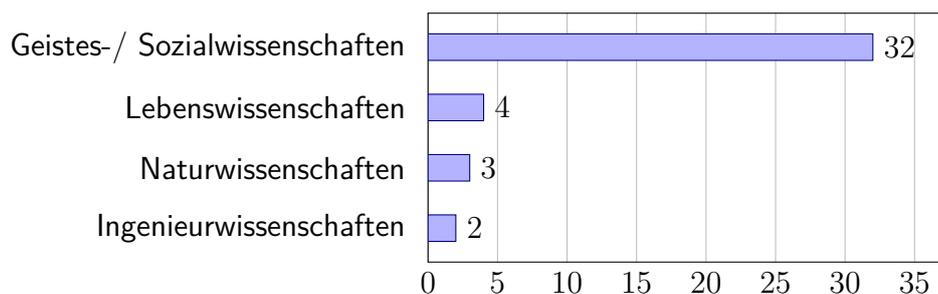


Abb. 22: Verteilung der User-Stories nach der Fachsystematik der DFG

4.2.2 Bedarf der Wissenschaftsgemeinschaft

Unter den 41 User-Stories in der Kategorie „Digitalisierung“ sind 16 von den jeweiligen Bearbeitern mehrfach, d. h. zwei- bis viermal, eingeordnet. Allerdings ist die Schnittmenge mit den anderen Kategorien so breit gestreut und dadurch auch unauffällig, dass dieser Aspekt kaum ins Gewicht fällt und im Folgenden unberücksichtigt bleibt.

Die User-Stories in der Kategorie „Digitalisierung“ lassen sich wiederum in zwei bzw. drei Untergruppen einteilen: die Digitalisierung von Quellenmaterial und Forschungsliteratur durch die zuständige bzw. besitzende Institution sowie die Digitalisierung von Forschungsdaten und Quellenmaterial im Rahmen der eigenen Forschung. Unter „Quellenmaterial“ werden nicht nur Dokumente wie Handschriften, Akten und Drucke, sondern auch nicht-textuelle Objekte wie historische Gegenstände aus Sammlungen sowie Audio- und Videoaufzeichnungen verstanden. Auf die Einzelheiten innerhalb der User-Stories wird in den folgenden Abschnitten eingegangen.

4.2.2.1 Digitalisierung von Quellenmaterial (21)

„Als Forschende, wünschen wir uns eine vorangetriebene (qualitative) Digitalisierung in Archiven, Sammlungen und Bibliotheken, um, mittels einfachen Zugangs, zu verarbeitende Forschungsdaten leicht nutzen zu können.“

Der auffälligste Wunsch der Forschenden ist, ganz allgemein, eine verstärkt vorangetriebene Digitalisierung von unterschiedlichem Quellenmaterial der Universitätsbibliotheken und verschiedener außeruniversitärer Institutionen, seien es Archive, Sammlungen oder Behörden. Hierunter fallen die klassischen Digitalisate älterer Textzeugen (Handschriften und Drucke), aber auch statistische Ergebnisdaten von Messungen und Erhebungen. Allein acht User-Stories kommen aus der Geschichtswissenschaft sowie vier aus den Philologien, also aus der Sprach- und Literaturwissenschaft.

In einigen User-Stories wird explizit von einer qualitativen Digitalisierung gesprochen. Darunter verstehen die Forschenden ausführliche (und standardisierte) Metadaten zu den Digitalisaten und die Möglichkeit einer maschinellen Weiterverarbeitung durch geeignete Dateiformate der digitalen Volltexte. Es ist zu beachten, dass der Wunsch nach einer verstärkt vorangetriebenen Digitalisierung nichts über die Qualität der Digitalisate aussagt und im Zweifel die Forschenden generell eine qualitative Digitalisierung der Objekte einer einfacheren und schnelleren vorzögen. Ein weiterer Aspekt, der in einigen User-Stories mal mehr, mal

weniger deutlich angesprochen wird, ist ein einfacher Zugang zu den Digitalisaten über das Internet, um den Aufwand der Beschaffung so gering wie möglich zu halten.

4.2.2.2 Eigene Digitalisierung und Finanzierungsmöglichkeiten (14)

„Als Forschende, wünschen wir uns Unterstützung an den Universitäten für die Digitalisierung von eigenen Forschungsdaten, um direkten Einfluss auf den Digitalisierungsprozess und somit auch auf die Qualität der Digitalisate nehmen zu können.“

In einigen User-Stories kommen Unterstützungserwartungen zum Ausdruck, beispielsweise bei der Digitalisierung von (selbst) untersuchtem Material – Text-, Bild-, analoges Audio- und Videomaterial sowie Mikrofilme – zu erhalten. In erster Linie hätten die Forschenden gern eine Dienstleistung an der eigenen Universität, die sie aktiv bei der digitalen Aufbereitung von Daten unterstützt oder ihnen diese Arbeit sogar abnimmt, wenn es sich um bestimmte Verfahren handelt, die ohne ein tieferes IT-Wissen oder den Einsatz bestimmter Geräte nicht selbst umgesetzt werden können. Außerdem wurden vereinzelt flexible Finanzierungsmöglichkeiten gewünscht, um in eigener Regie die zur Institution zugehörigen Sammlungen zu digitalisieren und für Interessierte bereitzustellen.

4.2.2.3 Bereitstellung von digitalisierter Sekundärliteratur (6)

„Als Forschende, wünschen wir uns eine vorangetriebene Digitalisierung der Bibliotheksbestände (unserer Universität) und eine Erweiterung des Angebots für E-Books, um einfacher auf Forschungsliteratur zurückgreifen und diese verarbeiten zu können.“

Gewissermaßen ein Sonderfall der Digitalisierung ist die Bereitstellung von forschungsrelevanter Sekundärliteratur als PDF oder E-Book. Der recht unspezifische Wunsch der Forschenden aus sehr verschiedenen Fachbereichen richtet sich in erster Linie an die Universitätsbibliotheken, die ihre eigenen gedruckten Bestände als Volltexte digitalisieren bzw. in Absprache mit Verlagen den Zugang zu digitalen Versionen aktueller Forschungsliteratur ermöglichen sollen. Allen User-Stories gemein ist, dass sich die Forschenden durch die Digitalisierung einen leichteren Zugang zu Forschungsliteratur und eine bequemere Arbeit mit dieser erhoffen.

4.2.3 Status quo

Den Status quo zur Digitalisierung kann man nur schwer fassen, weil sie weltweit von unzähligen Institutionen betrieben wird. Archive, Sammlungen und Bibliotheken betreiben die Digitalisierung weitgehend unabhängig und in eigener Regie. In Deutschland unterstützt die DFG seit nunmehr fast 20 Jahren die Digitalisierung forschungsrelevanter Bestände in verschiedenen Programmen und hat Empfehlungen und Richtlinien zur Digitalisierung⁸⁴ entwickelt, die für Förderempfänger verbindlich sind. Außerdem unterstützt die DFG mehrere Projekte, die die Sammlung und Bereitstellung von Daten oder Texten zu oder aus einer Epoche bzw. eines Themengebiets über ein zentrales Webportal anstreben. Im Rahmen einer

⁸⁴ DFG-Praxisregeln „Digitalisierung“ (02/2013): http://www.dfg.de/formulare/12_151/12_151_de.pdf [letzter Zugriff 9.5.2015].

Public-Private-Partnership mit Google digitalisiert die Bayerische Staatsbibliothek (BSB) ihren gesamten urheberrechtsfreien historischen Bestand an Druckwerken. Nach Beendigung des Projektes sollen mehr als eine Million Titel digitalisiert vorliegen.⁸⁵ Ein Teil dieser digitalisierten Bestände der BSB steht bereits frei online zur Verfügung und soll auch in nachfolgenden DFG-Projekten zur Textaufbereitung einbezogen werden.

Alle Hochschulstandorte in Baden-Württemberg bieten über ihre Webseiten prinzipiell die Möglichkeit, Digitalisate anzusehen und herunterzuladen.⁸⁶ Natürlich variiert der Umfang der angebotenen Digitalisate von Standort zu Standort mitunter beträchtlich. Seit 2001 werden fünf Universitäts- und Landesbibliotheken mit umfangreichen historischen Sammlungen bei der Digitalisierung mit Landesmitteln unterstützt. In der eigentlichen Digitalisierung und insbesondere in der Digitalisierung von historischen Texten nimmt die Universität Heidelberg eine hervorgehobene Stellung ein, weil auf ihre Software und ihren Workflow *DWork*⁸⁷ nicht nur die Digitalisierungsstellen der baden-württembergischen Universitätsbibliotheken Freiburg und Tübingen, sondern auch weitere nationale und internationale Einrichtungen zurückgreifen.

Die Vernetzung der dezentralen Digitalisierung erfolgt oft über fachspezifische Portale, in denen die wichtigsten bibliographischen Informationen zu den aufgenommenen Digitalisaten gesammelt sind. Inwiefern diese Portale den Forschenden bzw. der jeweiligen Community bekannt sind und von ihnen auch genutzt werden, konnte nicht festgestellt werden. Beispielhaft sollen hier einige nationale Projekte genannt werden.

Für die Katalogisierung von Handschriften und gegebenenfalls die Verlinkungen zu den dazugehörigen Digitalisaten zeigt das von der DFG unterstützte Großprojekt *Manuscripta mediaevalia*⁸⁸, das sich auf die mittelalterlichen Handschriften im deutschsprachigen Raum konzentriert, wie über ein einheitliches Webportal die Digitalisate der verschiedenen Digitalisierungsstellen – hier in erster Linie Bibliotheken und Archive – über eine Webseite abrufbar gemacht werden können, obwohl die Digitalisate dezentral von den Eigentümern angeboten werden. Durch *Manuscripta mediaevalia* wird den Forschenden die Möglichkeit gegeben, ortsunabhängig, kostenlos und zentral im mittelalterlichen Handschriftenbestand zahlreicher Institutionen im deutschsprachigen Raum zu suchen.

⁸⁵ Vgl. <https://www.bsb-muenchen.de/die-bayerische-staatsbibliothek/projekte/digitalisierung/massendigitalisierung-im-rahmen-einer-public-private-partnership-zwischen-der-bayerischen-staatsbibliothek-und-google/> [letzter Zugriff 10.5.2015].

⁸⁶ Es sind: Freiburg (<http://www.ub.uni-freiburg.de/index.php?id=dipro>), Heidelberg (<http://www.ub.uni-heidelberg.de/helios/digi/digilit.html>), Hohenheim (<https://kim.uni-hohenheim.de/100410>), Karlsruhe (<http://www.bibliothek.kit.edu/cms/678.php>), Konstanz (<http://www.ub.uni-konstanz.de/digitale-bibliothek/>), Mannheim (<http://www.bib.uni-mannheim.de/361.html?&L=0>), Stuttgart (<https://www.ub.uni-stuttgart.de/suchen-bestellen/e-ressourcen/index.html>), Tübingen (<http://idb.ub.uni-tuebingen.de/digitue/tue/>) und Ulm (<https://www.uni-ulm.de/einrichtungen/kiz/bibliothek/digitale-bibliothek.html>).

⁸⁷ Siehe dazu: http://www.ub.uni-heidelberg.de/helios/digi/tech_workflow.html und <http://www.ub.uni-heidelberg.de/helios/digi/dwork.html>; [letzter Zugriff 9.5.2015].

⁸⁸ Homepage des Projektes: <http://www.manuscripta-mediaevalia.de> [letzter Zugriff 10.5.2015].

Im *Deutschen Textarchiv* (DTA)⁸⁹ werden kostenfrei über 1600 Volltexte von ca. 1600 bis 1900 angeboten. Der Nutzer des von der DFG geförderten DTA kann diese Texte und / oder ihre Metadaten in verschiedenen Formaten herunterladen. Der Volltext ist als XML / TEI, HTML und einfache Textdatei abrufbar. Die Metadaten liegen als Dublin Core, CDMI und TEI-Header vor. Darüber hinaus wurde mit einem *Tokenizer*⁹⁰ in allen Texten eine Wortartenbestimmung aller vorkommenden Wörter durchgeführt.

Ein weiteres Großprojekt mit zentralem Webportal ist die *Deutsche Digitale Bibliothek* (DDB)⁹¹, in der der Nutzer in den digitalisierten Beständen samt ihren Erschließungsinformationen aus Kultur- und Wissenschaftseinrichtungen wie Bibliotheken, Archiven, Museen, Denkmalämtern, Mediatheken sowie Universitäten und anderen Forschungseinrichtungen suchen kann. Der Bestand umfasst digitale Objekte, die nicht gleichbedeutend mit digitalem Text zu verstehen sind, auch wenn dieser Anteil sicherlich der größte ist. Als ambitioniertes Ziel setzt sich die frei zugängliche DDB „die digitalen Angebote der deutschen Kultur- und Wissenschaftseinrichtungen sukzessive miteinander zu verlinken und zugänglich zu machen.“⁹² Die DDB soll in die europäische virtuelle Bibliothek *Europeana*⁹³ integriert werden.

Außerdem existieren fachspezifische oder thematische Webportale, die eine bestimmte wissenschaftliche Klientel bzw. Community ansprechen sollen. In den ebenfalls durch die DFG geförderten virtuellen Fachbibliotheken (ViFa) und virtuellen Fachportalen⁹⁴ können die Forschenden nach Fachrichtungen gefilterte Informationen abrufen. Hier versammeln sich Metadaten zu nicht digitalisierten Objekten und digitale bzw. digitalisierte Objekte.

Zuletzt soll hier auf die *Elektronische Zeitschriftenbibliothek* (EZB)⁹⁵ hingewiesen sein. Sie bietet den Forschenden an Universitäten und anderen Einrichtungen prinzipiell die Möglichkeit, online und für den Forscher kostenlos auf Aufsätze als Digitalisate und zum Teil auch als digitalisierte Volltexte zuzugreifen, ganz wie es sich die Forschenden in [4.2.2.3. Bereitstellung von digitalisierter Sekundärliteratur](#) wünschen. Jedoch variieren an den einzelnen Hochschulstandorten in Baden-Württemberg die Zugriffsrechte auf den Gesamtbestand der EZB von Standort zu Standort aufgrund der eingekauften Lizenzen. Der „freie“ Zugang der Forschenden auf digitalisierte Literatur ist also durch die Lizenzen an einer Universität mitunter stark eingeschränkt.

Häufig werden in Forschungsprojekten, die auf historische Quellen zurückgreifen, diese intern derart aufgearbeitet, dass es nur noch wenig zusätzlichen Aufwands bedarf, um diese Quellen qualitativ zu digitalisieren, weil der maschinenlesbare Text und / oder die Metadatenanreicherung zum Text (nach anerkannten Standards⁹⁶) durch die Forschenden zu großen

⁸⁹ Homepage des Projektes: <http://deutschestextarchiv.de> [letzter Zugriff 10.5.2015].

⁹⁰ Eine Software zur automatisierten Wortartenbestimmung. Vgl. <https://de.wikipedia.org/wiki/Tokenizer> [letzter Zugriff 10.5.2015].

⁹¹ Homepage des Projektes: <https://www.deutsche-digitale-bibliothek.de> [letzter Zugriff 10.5.2015].

⁹² Siehe <https://www.deutsche-digitale-bibliothek.de/content/faq#188> [letzter Zugriff 10.5.2015].

⁹³ Homepage des Projektes: <http://www.europeana.eu/portal> [letzter Zugriff 10.5.2015].

⁹⁴ Eine Übersicht findet sich hier: <http://blogs.sub.uni-hamburg.de/webis/fachportale> [letzter Zugriff 10.5.2015].

⁹⁵ Homepage des Projektes: <http://ezb.uni-regensburg.de> [letzter Zugriff 10.5.2015].

⁹⁶ Vgl. [5.1.2.6. Standardisiertes Metadatenschema](#).

Teilen bereits bearbeitet wurden. Solche Vorarbeiten werden meistens nicht publiziert und sind dadurch innerhalb der Community kaum bekannt, so dass sie am Ende in Vergessenheit geraten und gegebenenfalls von anderen erneut durchgeführt werden.

4.2.4 Gegenwärtige Handlungsempfehlungen

Eine erste wichtige Erkenntnis der untersuchten User-Stories ist eine recht große Zufriedenheit mit der Digitalisierung von textlichen Quellen durch deren heutige Besitzer (Archive, Sammlungen, Bibliotheken, Museen etc.), denn in den Wünschen war immer wieder zu lesen, dass die Digitalisierung verstärkt oder in qualitativer Form vorangetrieben werden sollte. Dies ist auch deswegen sehr erfreulich, weil es nur wenige Möglichkeiten gäbe, von außen Einfluss auf den Digitalisierungswillen dieser Einrichtungen zu nehmen. Außerdem kommt hinzu, dass die Einrichtungen, die über forschungsrelevante Daten verfügen, weltweit verteilt sind, was die Abstimmung in vielerlei Hinsicht erschwert.

Weil die Forschenden Qualität an die Aspekte Maschinenlesbarkeit, Metadatenanreicherung und bisweilen sehr konkrete Vorstellungen knüpfen, ergibt sich daraus der logische Schluss, die Zusammenarbeit zwischen Digitalisierungsprojekten bzw. Besitzern des Quellenmaterials, und Spezialisten bzw. interessierten Forschenden, zu lancieren bzw. zu verbessern, um die Digitalisate für die Forschung gewinnbringend nutzbar zu machen. Insbesondere für die Erstellung von maschinenlesbaren Texten von Quellenmaterial in fremden Sprachen und Schriften, bei denen die Anwendung von gewöhnlichen, automatisierten OCR-Verfahren nicht sinnvoll ist, erscheint eine Zusammenarbeit mit Spezialisten für eine qualitative Digitalisierung eines Textes erfolgsversprechend.

Forschungsorientierte Digitalisierung Durch eine stärkere Kopplung der Digitalisierung an aktuelle Forschung sollte es ebenso möglich sein, Forschungsvorarbeiten in qualitative Digitalisate umzusetzen. Es wäre zu begrüßen, wenn die Finanzierung einer speziellen Digitalisierung im Rahmen von Forschungsprojekten durch entsprechende Programme ermöglicht würde. Denkbar ist dabei eine verpflichtende Kooperation mit einer der Digitalisierungsstellen der baden-württembergischen Universitäten, um den Digitalisierungsprozess durch entsprechende Fachleute begleiten zu lassen.

Reihenfolge der Digitalisierung Wenn Bibliotheken ohnehin ihre Bestände digitalisieren, geht es hier nur um die Frage der Reihenfolge, damit die Forschenden das, was sie aktuell brauchen, digital zur Verfügung haben. Eine Lösung könnte ein einfach zugängliches System zur allgemeinen „*Digitalisierung on demand*“ sein, das Forschenden die Möglichkeit böte zu äußern, welche Werke digitalisiert werden sollen.

Fernleihe digitaler Inhalte Eine zufriedenstellende Umsetzung einer „Leihe“ von digitalisierter Literatur ist nicht leicht zu erzielen, weil verschiedene rechtliche Aspekte im Sinne eines Lizenzvertrages nicht geklärt sind.⁹⁷ Auf einen kleinen Nutzerkreis beschränkte, digitale

⁹⁷ Vortrag von Harald Müller während des E-Book-Workshops für MPG-Bibliotheken (Leipzig 2012).

Inwiefern sich eine „Leihe“ von Digitalisaten gegenüber einem Kauf unterscheidet und ob Digitalisate von

Lesesäle, bieten nicht den Komfort, den sich die Forschenden wünschen und führen die digitalen Möglichkeiten ad absurdum. Hier sollte in Zusammenarbeit mit den Verlagen eine für alle Beteiligten zufriedenstellende Lösung gefunden werden. Den Bibliotheken bietet Streaming sehr wahrscheinlich einen Ausweg aus den aktuellen Schwierigkeiten der digitalen Buchleihe.

Technische Weiterentwicklung Um eine an der Forschung ausgerichtete qualitative Digitalisierung umsetzen zu können, bedarf es des Ausbaus der derzeitigen Infrastruktur und der Weiterentwicklung der Digitalisierungssoftware (OCR), um auf die verschiedenen Bedürfnisse in der Forschung und in den einzelnen Forschungsprojekten eingehen zu können. Zudem kann der bisherige Digitalisierungsworkflow der Digitalisierungsstellen schrittweise derart umgebaut werden, dass diese auch die Zusammenarbeit mit Fachspezialisten und einzelnen Forschungsprojekten koordinieren und so den Erfolg eines von den Forschenden angestoßenen Digitalisierungsprojektes sicherstellenn.

4.3 Wissenschaftliche Zusammenarbeit (145)

4.3.1 Einleitung

Informationstechnische Dienste und neue Kommunikationstechnologien bieten eine Vielzahl an Möglichkeiten zur Unterstützung wissenschaftlicher Zusammenarbeit, was sich wiederum in einer zunehmenden Nutzung solcher Dienste in der alltäglichen Arbeit von Forschenden widerspiegelt. Die Manifestation von Wissen, Information und Forschungsgegenstand in elektronischen Daten und – damit einhergehend – der Austausch derselben, spielt eine immer größere Rolle. Dies beginnt schon auf der Ebene der Arbeitsgruppe, innerhalb derer gemeinsam Primär- und Sekundärdaten bearbeitet werden, und führt weiter über Instituts- und Ländergrenzen hinweg, da auch der Austausch von Informationen und Daten mit externen bzw. internationalen Partnern immer weiter an Bedeutung gewinnt. Dies ist beispielsweise in den Lebenswissenschaften der Fall, wenn Forschung auf spezielle ortsgebundene Infrastruktur aufbaut und Daten experimenteller Ergebnisse an anderen Standorten weiterverarbeitet werden. Auch in den Geistes- und Gesellschaftswissenschaften beflügeln neue technische Möglichkeiten die Zusammenarbeit und die Entwicklung neuer Forschungsansätze. Um eine solche wissenschaftliche Kooperation zu fördern, ist es wichtig, benutzerfreundliche und zuverlässige Dienste sowie eine IT-Infrastruktur bereitzustellen, die eine ortsunabhängige Kollaboration ermöglicht. Weiterhin stellt Datensicherheit für wissenschaftliche Kooperation einen zentralen Punkt dar, der in Bereichen personenbezogener Forschung besondere Bedeutung hat. Schließlich kommt es in Bezug auf die Verwaltung der Daten auf Benutzbarkeit an, die einerseits auf Seiten der Forschenden Einfachheit erfordert und andererseits – beispielsweise unterstützt durch Metadaten oder Versionierung – den Anforderungen an die nachvollziehbare Verständlichkeit der Daten genügt.

Text wie Bücher oder eher wie Software behandelt werden müssen, ist nicht genau definiert und somit bisher rechtlich ungeklärt.

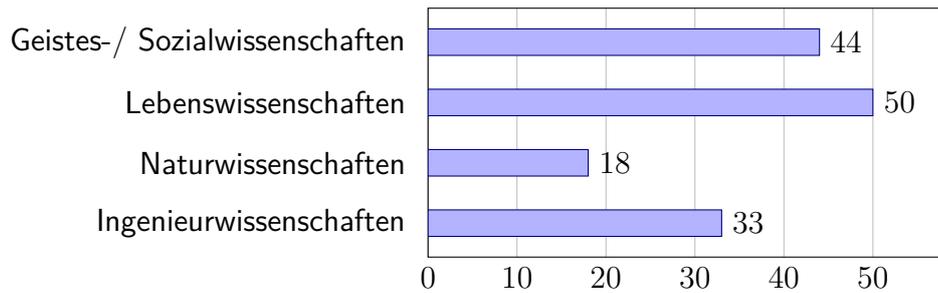


Abb. 23: Verteilung der User-Stories nach der Fachsystematik der DFG

4.3.2 Bedarf der Wissenschaftsgemeinschaft

Im Folgenden werden zunächst die Antworten auf die konkreten Interviewfragen beschrieben, die sich auf den Austausch von Forschungsdaten beziehen. Im Anschluss folgen die Bedürfnisse der befragten Forschenden anhand zusammengefasster User-Stories.

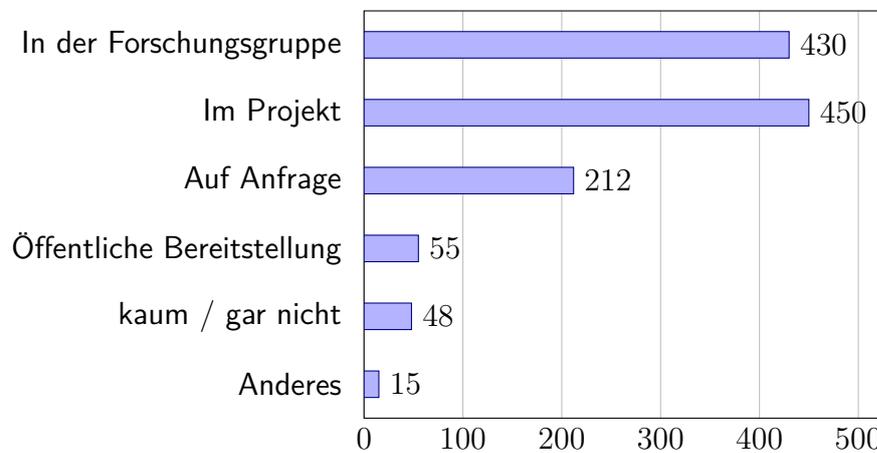


Abb. 24: „In welchem Rahmen haben Sie schon eigene Primärdaten ausgetauscht?“ (Frage 20a – Mehrfachnennung der vorgegebenen Austauschvarianten möglich)

Auf die Frage, in welchem Rahmen bereits Primärdaten ausgetauscht wurden, gaben die meisten Befragten an, innerhalb der Forschungsgruppe oder innerhalb eines Projekts Primärdaten ausgetauscht zu haben (s. Abb. 24). Der Austausch von Primärdaten mit externen Personen erfolgte bei 212 von 571 Befragten nur auf Anfrage. Lediglich 55 der 571 Personen gaben an, ihre Daten öffentlich zum Download bereitzustellen. Zwischen den einzelnen Fachbereichen ergaben sich im Anteil der jeweils Befragten keine wesentlichen Unterschiede. Die Antworten legen nahe, dass der Austausch von primären Forschungsdaten einen wesentlichen Aspekt in der Zusammenarbeit innerhalb einer Forschungsgruppe bzw. innerhalb eines Forschungsprojektes darstellt. Trotzdem wird ein öffentlicher Austausch dieser Daten häufig nicht betrieben.

Auf die Frage, in welchem Zusammenhang und mit welchen Programmen oder Diensten Forschungsdaten ausgetauscht werden, wurden einige Angebote genannt, die in Abb. 25 mit ihrer Nennungshäufigkeit aufgelistet sind. Der meist gewählte Weg ist der Datenaustausch

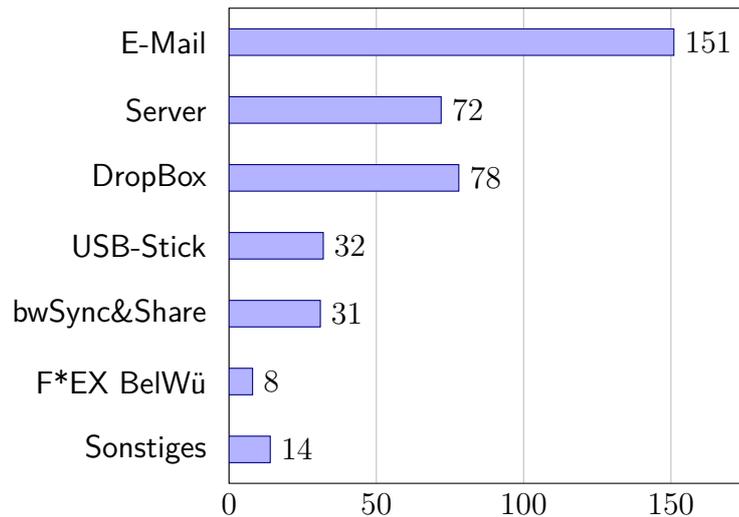


Abb. 25: „Wenn Austausch: In welchem Zusammenhang? Welche Programme oder Dienste nutzen Sie dafür?“ (Frage 20b – Mehrfachnennung einzelner Angebote möglich)

per E-Mail mit 151 Erwähnungen, gefolgt von DropBox mit 78 und der Austausch über einen Server mit 72 Nennungen. Der erst seit Beginn des Befragungsjahres zur Verfügung stehende Landesdienst bwSync&Share⁹⁸ wurde 31 mal genannt. F*EX⁹⁹, als der Datentransferdienst des BelWü Netzwerks wurde 8 mal angesprochen.

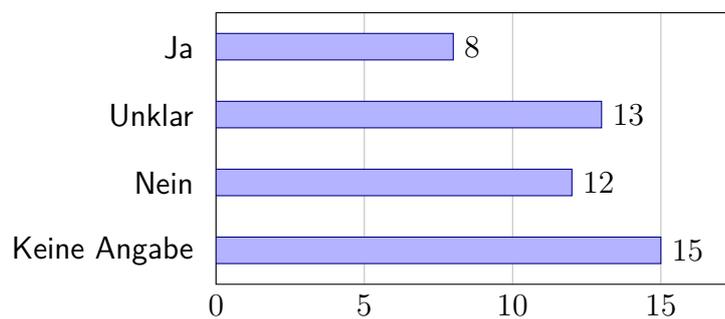


Abb. 26: „Wenn kaum Austausch: Gibt es Partner oder Kollegen mit denen Sie den Informationsaustausch gern verbessern würden?“ (Frage 20c)

Aufbauend auf die vorangegangenen Fragen wurden diejenigen, die angaben kaum oder gar nicht Primärdaten ausgetauscht zu haben, danach befragt, ob es Kollegen oder Partner gibt, mit denen sie den Austausch verbessern wollen würden. Die Antworten auf diese Frage wurden in „ja“, „unklar“, „nein“ und „keine Angabe“ (s. Abb. 26) kategorisiert. Hier sind aber nur diejenigen Befragten eingeschlossen, die von sich selbst gesagt haben, dass sie kaum oder gar nicht austauschen.¹⁰⁰ Ob weitere Angaben (beispielsweise „Anderes: Koautoren“)

⁹⁸ bwSync&Share, <http://www.alwr-bw.de/kooperationen/bwsync-share/> [letzter Zugriff 28.3.2015].

⁹⁹ F*EX, <http://fex.belwue.de/index.html> [letzter Zugriff 28.3.2015].

¹⁰⁰ In beliebigen anderen Antwortkombinationen.

vielleicht auch der Kategorie „selten bis nie“ zufallen würde, wurde hier nicht weiter bewertet. Von den dann noch klar als „ja, ich möchte verbessern“ klassifizierten Forschenden (8) gaben fünf an, dass sie eine technische Lösung brauchen. In den „unklar“ kategorisierten Antworten stehen immer Antworten auf Frage 20c, die kein direktes, sondern bestenfalls ein implizites „ja“ bedeuten. Das häufigste Beispiel ist „Die Daten sind zu groß.“ Das ist nicht genau eine Antwort auf die Frage, ob sie den Austausch verbessern wollen, lässt aber womöglich indirekt auf den Wunsch nach einer technischen Lösung schließen. Dem gegenüber stehen die Befragten die hier „nein“ sagten. Diese waren ihrer Angabe zufolge meist die Einzigen die die Daten nutzen wollen oder können. An dieser Frage zeigt sich exemplarisch wie viel Detailarbeit in die Analyse der Antworten gesteckt werden kann, um die Bedarfe zu analysieren. Wir halten hier aber nur noch für den Großteil der Befragten fest (auch die, die Austausch betreiben), dass eine gute technische Lösung für der Mehrheit der Befragten eine Relevanz hätte, wobei bwSync&Share teilweise auch schon als Verbesserung oder gar Lösung angesehen wurde.

Im Folgenden werden nun die sich aus den Gesprächen ergebenden User-Stories zur wissenschaftlichen Zusammenarbeit dargestellt.

4.3.2.1 Interner Datenaustausch (56)

„Als Forschende, die große Daten schnell und einfach mit Kollegen vor Ort, anderen Fachkollegen und internationalen Partnern austauschen möchten, wünschen wir uns einen stabilen, benutzerfreundlichen, sicheren und schnellen Datenaustauschdienst, um den Datenaustausch einfacher und gleichzeitig sicher zu gestalten und somit die Zusammenarbeit zu verbessern.“

Über alle Fachbereiche hinweg besteht der Wunsch nach einem einfachen, zentralisierten Datenaustauschdienst, mit dessen Hilfe Daten – auch mit internationalen Partnern – sicher ausgetauscht werden können. Außerdem besteht der Wunsch nach größeren E-Mail-Speichern zum Versenden großer Anhänge. Besonders in den Naturwissenschaften stellt der Austausch sehr großer Datenmengen eine Herausforderung für die Forschenden dar. Der Landesdienst bwSync&Share soll nach Wunsch der Forschenden mehr Speicherkapazität bieten, stabiler und benutzerfreundlicher werden (beispielsweise durch eine Sprachauswahl) und verschlüsselt sein, um eine höhere Datensicherheit zu gewährleisten. Hier sei angemerkt, dass zum Zeitpunkt der Befragungen der Speicherplatz pro Anwender auf 10 GB begrenzt war und inzwischen auf 25 GB erhöht wurde. Manche Befragte gaben an, bwSync&Share kurz nach dem Erscheinen getestet und nach den ersten Downtimes jedoch nicht weiter verwendet zu haben. Im Hinblick auf Datensicherheit haben Forschende, die mit personenbezogenen Daten arbeiten, einen besonderen Bedarf. Dies betrifft besonders die medizinisch-naturwissenschaftlichen Fächer, die sensible Daten in großer Menge mit externen Einrichtungen wie Kliniken und Ämtern austauschen müssen.

4.3.2.2 Interne Datenverwaltung (56)

„Als Forschende, die größere Datenmengen mit Projektpartnern kollaborativ verarbeiten, wünschen wir uns eine sichere und zentrale Speicherlösung mit optionaler Metadatenstruktur und Versionsverwaltung sowie flexibler Rechtevergabe und Schnittstellen zu weiterer IT-“

Infrastruktur (u.a. HPC), um ortsunabhängig mit übersichtlichen und datenschutzkonformen Systemen arbeiten zu können.“

Angesichts datenintensiver Forschung, Projektpartnern an unterschiedlichen Orten und Fluktuationen in den Forschungsgruppen, wünschen sich die Forschenden zentral bereitgestellte Speicherlösungen, um potentiell große Datenmengen ablegen und für eine ortsunabhängige Bearbeitung verfügbar halten zu können. Um weiterhin die Mobilität der Forschenden zu unterstützen, wünschen sich diese konfigurierbare Cloud-Lösungen für ihre Forschungsgruppen, um interne Daten schneller verfügbar und zentral abrufbar zu halten. Die Datenverwaltung soll zentral durch das Land bzw. die Rechenzentren in Form von Richtlinien und Verwaltungshilfen unterstützt werden und das Aufsetzen von Projektdatenbanken soll erleichtert werden. Für die Verwaltung ihrer Daten wünschen sich die Forschenden eine flexible Zugriffsrechtevergabe. Außerdem sollen durch Metadatenstrukturen und Möglichkeiten der Versionierung transparente Datenstrukturen geschaffen und Redundanzen vermieden werden. Forschende, die mit personenbezogenen Daten arbeiten, haben einen besonderen Bedarf nach erhöhten (zertifizierten) Sicherheitsstandards bei der Verwaltung ihrer Daten.

4.3.2.3 Gemeinsame und gleichzeitige Datenbearbeitung (15)

„Als Forschende, die eng mit Projektpartnern zusammenarbeiten, wünschen wir uns eine online-Plattform zur Bearbeitung von Daten und Textdokumenten mit integrierter Versionsverwaltung, um mobil auf Dokumente und Daten zugreifen und diese gleichzeitig mit Projektpartnern bearbeiten zu können.“

Bedingt durch die enge Kooperationen mit Kollegen sowie die vermehrte Interdisziplinarität und Internationalität in den Forschungsgruppen besteht der Wunsch nach einer webbasierten Plattform, die die gleichzeitige Bearbeitung von Dokumenten und Daten erlaubt und über eine Versionsverwaltung verfügt, um ggf. ältere Dateiversionen wiederherstellen zu können. Darauf aufbauend besteht der Wunsch, mobil per Netzzugriff auf bereits gespeicherte Daten und Dokumente zugreifen zu können, die dann einfach mobil bearbeitet und mit den Projektpartnern geteilt werden können. Die Forschenden könnten somit das als umständlich empfundene Versenden großer Datenmengen per E-Mail umgehen. Solche Dienste würden als Verbesserung der Zusammenarbeit und ferner als Verbesserung der virtuellen Arbeitsumgebung angesehen werden. In den User-Stories dieser Kategorie zeigt sich eine Tendenz für solche Bedarfe besonders in den Gesellschafts- und Geisteswissenschaften (etwa zwei Drittel der User-Stories). Durch neue Geräte die „virtuelle Realität“ versprechen, werden in wenigen Jahren leistungsstarke und mobile Endgeräte zur Verfügung stehen, die diesen Bedarf in neuer Form noch einmal massiv verstärken könnten.

4.3.2.4 Groupware (19)

„Als Forschende, die internationale Kooperationen pflegen und institutionenübergreifend Projekte koordinieren, wünschen wir uns effiziente Tools für Webkonferenzen, Projektverwaltung und standardisierte Groupware, um gemeinsame Projekte einfacher planen und durchführen zu können und den (internationalen) Informationsaustausch mit Kollegen zu erleichtern.“

Angesichts einer Vielzahl an E-Mail-Programmen, bei denen die Kalender nicht anwendungsübergreifend geteilt werden können, und kommerzieller Kommunikationssoftware (Skype etc.) besteht der Wunsch nach einheitlichen, zentral angebotenen Diensten, die das Planen und Koordinieren von Projekten erlauben, wie beispielsweise eine verbesserte Aufgabenteilung durch einen gemeinsamen Kalender und gemeinsame Kontaktverwaltung. Idealerweise können diese Programme mit den verschiedenen Endgeräten der Forschenden synchronisiert werden oder verfügen über Schnittstellen zu weiteren zentralen Diensten.

4.3.3 Status quo

Eine nähere Betrachtung von Frage 20 (Abb. 14 auf Seite 32) zeigt, dass innerhalb von Projekten und Forschungsgruppen Daten größtenteils durch Freigaben über das Netzwerk ausgetauscht werden. Falls es sich um kleinere Datenmengen handelt, werden des öfteren E-Mails oder bwSync&Share¹⁰¹ verwendet. In manchen Fachgebieten werden noch immer Daten auf Festplatten oder CDs per Post an ihr Ziel versendet, weil die Infrastruktur zwischen den beteiligten Instituten eine digitale Übertragung nicht in einer akzeptablen Zeit zulässt.

Der folgende Abschnitt gibt einen kurzen Überblick über einige vorhandene IT-Angebote im Land Baden-Württemberg, die die wissenschaftliche Zusammenarbeit unterstützen. Die Auswahl der hier vorgestellten Angebote wurde anhand der in den User-Stories erwähnten, bereits getesteten oder genutzten Dienste sowie den offenen Wünschen der Forschenden getroffen.

bwIDM Mit bwIDM¹⁰² wurde ein föderiertes Identitätsmanagement für die Hochschulen in Baden-Württemberg geschaffen. Es bildet eine integrale Grundlage für Authentifikation und Autorisation von Nutzern zum Zugriff und der Verwendung landesweiter Dienste wie bwSync&Share oder bwFileStorage¹⁰³. Dadurch können Landesdienste ortsunabhängig und sicher genutzt werden.

Datenaustauschdienste bwSync&Share ist ein Online-Speicherdienst, der das Hochladen und Teilen von Dokumenten zwischen verschiedenen Endgeräten und Nutzern ermöglicht. Die Daten werden über die Large-Scale-Data-Facility (LSDF) am KIT gesichert. Das Authentifizierungsmanagement ist mittels bwIDM realisiert. Anfang 2015 wurde der verfügbare Speicherplatz von 10 GB auf 25 GB erhöht. Es existieren Kostenmodelle, die eine über diesen Speicherplatz hinausgehende Nutzung des Dienstes ermöglichen. Durch das Herunterladen einer Client-Software besteht die Möglichkeit, von unterschiedlichen Endgeräten auf die eingestellten Inhalte zuzugreifen. Unterstützt werden die Betriebssysteme Windows, Mac OS, Linux, Apple iOS und Android. Die Kommunikation zwischen den Endgeräten und dem KIT erfolgt verschlüsselt. Die Ablage der Daten auf den Online-Speichersystemen am KIT ist

¹⁰¹ bwSync&Share, <https://bwsyncandshare.kit.edu/login> [letzter Zugriff 19.6.2015].

¹⁰² bwIDM, <https://www.bwidm.de/> [letzter Zugriff 11.5.2015].

¹⁰³ bwFileStorage, <http://www.alwr-bw.de/kooperationen/bwfilestorage/> [letzter Zugriff 28.3.2015].

unverschlüsselt. Aktuell wird eine Erweiterung zur Möglichkeit der gemeinsamen gleichzeitigen Dokumentbearbeitung entwickelt, die auf der Implementierung von Power-Folder basiert. Von dieser ausgehend wird es erheblichen Aufwand bedeuten, ein ebenbürtiges Produkt zu Google Docs zu entwickeln, allerdings lassen sich hier vielleicht Synergien nutzen, um auch gemeinsames bearbeiten anderer Datenarten zu ermöglichen. Innerhalb des Horizon-2020-Projektes EUDAT wird mit B2DROP¹⁰⁴ ein europäischer Datenaustauschdienst entwickelt. Dieser basiert auf ownCloud¹⁰⁵ und ist frei verwendbar für Forschende aus Europa. Über das Landeshochschulnetz BelWü wird der Service F*EX angeboten, der das Versenden von großen Dateien erlaubt. Der hypertextbasierte Dienst funktioniert ähnlich wie eine E-Mail, in der eine oder mehrere Empfangsadressen angegeben werden, die dann per Mail einen Download-Link erhalten. Es wird ein beliebiges E-Mail-Programm und ein Webbrowser benötigt. Eine Anmeldung bei F*EX ist nur für den Sender, nicht aber für den Empfänger nötig. Dazu wird eine E-Mail-Adresse aus dem BelWü-Netz benötigt, d.h. eine über BelWü oder an einer der Landesuniversitäten registrierte Domain.

bwFileStorage Über bwFileStorage haben Angehörige der Landesuniversitäten die Möglichkeit, Daten in höherem Umfang auf der LSDF am KIT zu speichern. Der Dienst stützt sich auf die Authentifizierung über bwIDM. Per Zugriffsrechtevergabe können weitere Personen auf die eingestellten Daten zugreifen. Der Speicherplatz wird anhand von Quotas zugeteilt. Für das Ablegen von personenbezogenen Daten ist die LSDF offiziell nicht geeignet.

bwMailService Seit 2014 läuft das Projekt bwMailService¹⁰⁶ mit der Zielsetzung, Rahmenbedingungen für eine kooperative Erbringung eines Mail- und Groupwaredienstes auszuarbeiten. Dazu sollen offene Standards unterstützt werden. Die Basis bildet dabei wieder die Anbindung an bwIDM. Die Ansätze zur Umsetzung sind zweierlei: Untersucht werden die Möglichkeit der Nutzung eines cloudbasierten Dienstes wie Microsoft Office 365 bzw. die Einrichtung einer Community-Cloud an den Hochschulen selbst.

EduPad Das EduPad¹⁰⁷ der Universität Konstanz bietet die Möglichkeit, einen Webspaces mit Versionsverwaltung einzurichten, in dem mehrere Nutzer gleichzeitig Textdokumente erstellen und bearbeiten können. Es können Dokumente im Format HTML, Word oder RTF importiert werden und Exporte als PDF, HTML, Textdatei, Word- oder OpenDocument-Format erfolgen. Es besteht ebenfalls die Möglichkeit ein passwortgeschütztes „ProPad“ einzurichten, das verschlüsselte Links nutzt. Bilder können hier nicht eingebunden werden und auch sonst liegt der Umfang der Funktionalität hinter dem von Google Docs.

¹⁰⁴ B2DROP, <http://eudat.eu/> [letzter Zugriff 28.3.2015].

¹⁰⁵ ownCloud, <https://owncloud.org/> [letzter Zugriff 28.3.2015].

¹⁰⁶ bwMailService, <http://www.alwr-bw.de/kooperationen/bwmailservice/> [letzter Zugriff 28.3.2015].

¹⁰⁷ EduPad, <http://www.rz.uni-konstanz.de/dienste/edupad/> [letzter Zugriff 28.3.2015].

BW-eLabs / BW-eSci(T) Das MWK förderte im Zeitraum von August 2009 bis Januar 2012 im Rahmen der E-Science-Strategie die Entwicklung des Pilotprojekts BW-eLabs¹⁰⁸. Als zentrale Disziplin wurde die Nanotechnologie ausgewählt, da hier der Zugang zu hochwertigen experimentellen Ressourcen aufgrund sehr hoher Kosten für alle Fachcommunitys einen großen Nutzen bringen würde. Digitale Bibliotheken und dezentrale Tools und Repositorien sind in die 3D-Plattform BW-eLabs eingebettet worden. Über ein zentrales Register sind verschiedene Wissenschaftseinrichtungen in einer Föderation vereint, innerhalb derer experimentelle Ressourcen per Buchungsdatenbank und Remote-Desktop-Funktion gebucht und gesteuert werden können. So wurden die zwei fernsteuerbaren Labore für Digitale Holografie und Synthese von Nanopartikeln entwickelt. Ebenfalls durch das MWK wurde das Projekt „BW-eSci(T) – Entwicklung einer prototypischen e-Science-Forschungsumgebung an der Universität Tübingen“¹⁰⁹ gefördert. Diese Umgebung soll über eine ausreichende Komplexität an Diensten, Benutzerschnittstellen und Datenmodellierung verfügen, so dass diese prototypische Forschungsumgebung zukünftig auf andere Fachbereiche übertragen werden kann. Das Projekt basiert auf der von MPG und FIZ Karlsruhe entwickelten Plattform und wird dementsprechend auch in enger Zusammenarbeit mit dem FIZ Karlsruhe geführt.

4.3.4 E-Science-Bestrebungen

Im ESFK werden Ansätze und Konzepte zur Stärkung der wissenschaftlichen Kollaboration im Rahmen von E-Science präsentiert. Durch die immer mehr ortsunabhängige, interdisziplinäre, internationale und kooperativ ausgerichtete Forschung erwachsen neue IT-Wünsche bei den Wissenschaften, die die Methodik der Wissenschaften nachhaltig verändern werden. Diesen Wünschen zu entsprechen, wird auch durch die DFG in ihrem Papier „Informationsverarbeitung an Hochschulen – Organisation, Dienste und Systeme“¹¹⁰ akzentuiert. IT-Dienstleistungen können nicht mehr singular ausgerichtet sein, sondern müssen der zunehmenden Kooperation in der Forschung Rechnung tragen. Konkreter spricht sich die DFG für die Entwicklung integrativer Plattformen aus, die die Arbeit der Forschenden unterstützen sollen:

„Benötigt werden Kooperationsplattformen für die Forscher der Hochschule und ihre weltweiten Partner, welche u. a. den Austausch von Informationen nach innen und außen, das Erstellen gemeinsamer Dokumente, das Führen gemeinsamer Kalender, die Planung von Projekten und die Verwaltung gemeinsamer Ressourcen erleichtern.“ (DFG, 2010)

In diesem Zusammenhang formuliert das MWK ein Förderprogramm zum Aufbau virtueller Forschungsumgebungen, die alle Schritte des wissenschaftlichen Wertschöpfungsprozesses

¹⁰⁸ BW-eLabs, http://www.bw-elabs.org/projekt/2_gesamtkonzept/index.html [letzter Zugriff 28.3.2015].

¹⁰⁹ BW-eSci(T), <http://www.uni-tuebingen.de/einrichtungen/zentrum-fuer-datenverarbeitung/projekte/bw-escit/projektuebersicht.html> [letzter Zugriff 28.3.2015].

¹¹⁰ Deutsche Forschungsgemeinschaft: Informationsverarbeitung an Hochschulen – Organisation, Dienste und Systeme, Empfehlungen der Kommission für IT-Infrastruktur für 2011–2015. Bonn, 2010.

unterstützen sollen. Dies beinhaltet die Zusammenführung zentraler Basisdienste, genauso wie die Integration fachspezifischer und disziplinübergreifender Dienste. Bedingt durch die Spezifität der in den einzelnen Disziplinen genutzten Analysewerkzeuge und der Breite an genutzten Datenformaten, empfiehlt das MWK den Aufbau solcher Kooperationsplattformen aus den Fachbereichen, die durch zentrale Infrastruktureinrichtungen proaktiv beraten und betreut werden sollen. Dies beinhaltet die Kofinanzierung nationaler und europäischer Projekte sowie Unterstützung bei der Überführung in den Dauerbetrieb. Im Rahmen der Strategie des MWK zur Ausweitung von konsortial erworbenen Lizenzen sollen Nutzungsformen und rechtliche Rahmenbedingungen für die Überführung in digitale Forschungsumgebungen geschaffen werden. Zusätzlich soll der Ausbau der technischen Infrastruktur nach den Bedarfen der Wissenschaft erfolgen. Das MWK will in enger Zusammenarbeit mit dem Bund, den Ländern und der DFG eine nachhaltig führende Rolle bei der Entwicklung digitaler Forschungsumgebungen einnehmen, die wissenschaftliche Kollaboration entscheidend verändern werden und befördern helfen sollen. Aufbauend auf den hier genannten Aspekten und den in [4.3.2. Bedarf der Wissenschaftsgemeinschaft](#) festgestellten Wünsche und Bedarfen der Forschenden sowie den vorhandenen Dienste im Land Baden-Württemberg werden im folgenden Abschnitt gegenwärtige Handlungsempfehlungen präsentiert, die zur weiteren Unterstützung und Verbesserung wissenschaftlicher Zusammenarbeit dienen sollen.

4.3.5 Gegenwärtige Handlungsempfehlungen

Die Auswertung der Interviews zeigt, dass die zunehmende Internationalisierung, institutionenübergreifende Kooperation und Mobilität die alltägliche Arbeit der Forschenden aller Disziplinen erheblich beeinflusst und eine Veränderung der wissenschaftlichen Methodik hin zur vermehrten Verwendung digitaler Ressourcen mit sich bringt. Daraus erwachsen für die Zukunft neue Bedarfe und Anforderungen an die IT-Infrastruktur an Baden-Württembergs Hochschulen.

Erweiterungskonzept der bisherigen Landesdienste zum Datenaustausch Die Befragung der Forschenden zeigt die Bereitschaft, den Landesdienst bwSync&Share zum Austausch von Daten zu nutzen. Die Erwartungen der Forschenden an den Dienst zielen auf Verbesserungen in Bereichen der Stabilität, Benutzerfreundlichkeit, Datenverschlüsselung und erhöhten Speicherplatz ab. Einer benutzerfreundlicheren Austauschmöglichkeit mit Forschenden aus anderen (Bundes-)Ländern könnte eine einfache Dokumentation dieses Features folgen, wobei auch auf eine notwendige englische Version des Dienstes vorhanden sein müsste. Für den Austausch großer Daten könnte die Integration einer Funktionalität gemäß F*EX in eine benutzerfreundliche Oberfläche für den allgemeinen Datenaustausch in Frage kommen. Des Weiteren könnte sich der Nutzerkreis durch eine Verschlüsselungsmöglichkeit erhöhen. In Bezug auf den Umgang mit und den Austausch von sensiblen und personenbezogenen Daten sei hier noch auf die Handlungsvorschläge in [2.1. Urheberrecht und Datenschutz](#) und [2.2. Informationsangebote](#) verwiesen.

Zentrale cloudbasierte Speicher, Repositorien und Datenbanken Forschende benötigen Unterstützung beim Aufsetzen und Betrieb von Datenbanken durch Einrichtungen wie etwa den Rechenzentren. Der Bedarf der Forschenden nach großen Speicherkapazitäten könnte zentral auf landesweiter Ebene adressiert werden. Der Impuls zu fachspezifischen Zusatzentwicklungen sollte dabei aus den Fachcommunitys kommen, die von einer einladenden Grundinfrastruktur profitieren wollen. Cloudbasierte Lösungen könnten als weiteres Angebot hin zu mehr Mobilität und Flexibilität für die Forschenden einerseits und der gemeinschaftlichen, kosteneffizienten Nutzung von IT-Ressourcen andererseits dienen¹¹¹.

Gleichzeitige Datenbearbeitung Forschende benötigen Arbeitsplattformen für eine gleichzeitige gemeinsame Bearbeitung von Textdokumenten oder Datensätzen. Die Befragten erhoffen sich durch zentrale Lösungen datenschutzrechtlich vorteilhafte Alternativen zu kommerziellen Plattformen. In Zusammenarbeit der Rechenzentren und Fachbereiche gilt es, Plattformen zu entwickeln, die neben der Textbearbeitung auch die Bearbeitung von Datensätzen ermöglichen. Eine Anbindung an verschiedene Dienste der Datenverarbeitung kann die wissenschaftliche Arbeit erleichtern und effiziente Workflows unterstützen.

Virtuelle Forschungsumgebungen Virtuelle Forschungsumgebungen können fachspezifisch aus den einzelnen Fachbereichen heraus entwickelt werden. Die Integration verschiedenster Dienste und Schnittstellen in zentralen Datenspeichern oder anderen physischen Ressourcen ist, vor allem in einigen Naturwissenschaften, nötig, um den wissenschaftlichen Schaffensprozess nachhaltig durch digitale Methoden zu unterstützen. Die einzelnen Infrastruktureinrichtungen können hierbei im Aufbau und bei der Bereitstellung generischer Hard- und Softwareinfrastruktur für virtuelle Forschungsumgebungen von zentraler Seite aus unterstützen.

¹¹¹ Etwaige Speicher- und Datenverwaltungslösungen stehen bereits im Fokus der Entwicklung im Landesprojekt bwCloud, <http://www.alwr-bw.de/kooperationen/bwcloud/> [letzter Zugriff 28.3.2015].

5 Informationsbewahrung

5.1 Projekt- und Datendokumentation (187)

5.1.1 Einleitung

Die begleitende Verwaltung und Dokumentation von Abläufen und Daten innerhalb eines Projektes wird mit der voranschreitenden Digitalisierung zunehmend wichtiger. Im Hinblick auf eine Langzeitarchivierung von Daten ist eine begleitende Dokumentation der Arbeitsschritte und des Lebenszyklus von Forschungsdaten notwendig. Vor diesem Hintergrund wurden die Forschenden gezielt nach ihrem Umgang mit diesen Anforderungen gefragt. Die Projekt- und Datendokumentation umfasst zum einen Wünsche nach Verbesserungen bei den Möglichkeiten der Verwaltung interner projektbezogener Daten im Sinne eines Projektmanagements. Zum anderen liegt ein Schwerpunkt auf der Verwaltung der Primärdaten mit einem besonderen Augenmerk auf deren Auffindbarkeit und Nachnutzung, sodass sich dieser Bereich im Wesentlichen auf die Anreicherung der Daten mit Metadaten konzentriert.

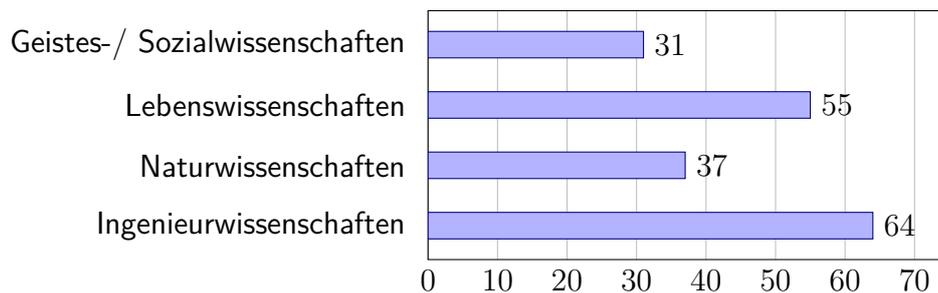


Abb. 27: Verteilung der User-Stories nach der Fachsystematik der DFG

Insgesamt wurden 187 User-Stories diesem Themenfeld zugewiesen. Die 64 User-Stories (34 %) aus den Ingenieurwissenschaften sind relativ zu ihrer Häufigkeit in der Stichprobe (23 %) ein hoher Wert, was für einen außerordentlich hohen Bedarf an Projekt- und insbesondere Datendokumentation in diesem Bereich hinweist.

5.1.2 Bedarf der Wissenschaftsgemeinschaft

Die Problematik bei der Projekt- und Datendokumentation lässt sich im Wesentlichen in zwei große Unterkategorien teilen:

- die Dokumentation interner Projektabläufe, die entweder aufgrund nicht ausreichender Beratung oder auch wegen Zeit- und Personalmangels häufig zu kurz kommt und in den

meisten Fällen frühestens am Ende des Projektes in das Blickfeld der Forschenden rückt – zu einem Zeitpunkt, an dem finanzielle und personelle Ressourcen bereits anderweitig, beispielsweise zur Planung eines Folgeprojekts, eingesetzt werden.

- die Anreicherung der Rohdaten mit Metadaten, die hinsichtlich der Anforderungen einer endgültigen Archivierung und / oder einer Nachnutzung erhobener bzw. erzeugter Daten grundlegend sind.

Projektdokumentation (105) Besonders bei der Verwaltung und Dokumentation der Projekte und der anfallenden Daten gibt es großen Unterstützungsbedarf, um die Forschenden bei diesen Aufgaben zu entlasten. Die Anforderungen und damit die Art der Unterstützung sind dabei sehr unterschiedlich. Zum einen fehlen beratende Stellen und Empfehlungen, wie ein effizientes und transparentes Projektmanagement aussehen kann, da diese Aspekte während des Studiums kaum angesprochen werden. Zum anderen mangelt es an Hilfen, die eine konkrete Umsetzung in der Praxis und eine Einbettung in den Forschungsalltag begleiten. Hierbei stellen sich die Befragten am ehesten die Entlastung durch automatisierte Prozesse in Form von Softwarelösungen vor oder aber den Einsatz zusätzlicher Personalmittel zur Auslagerung dieser Tätigkeiten an Fach- bzw. Hilfskräfte. Die insgesamt 105 User-Stories, die hierunter fallen, lassen sich demnach gemäß [Abb. 28](#) gruppieren.

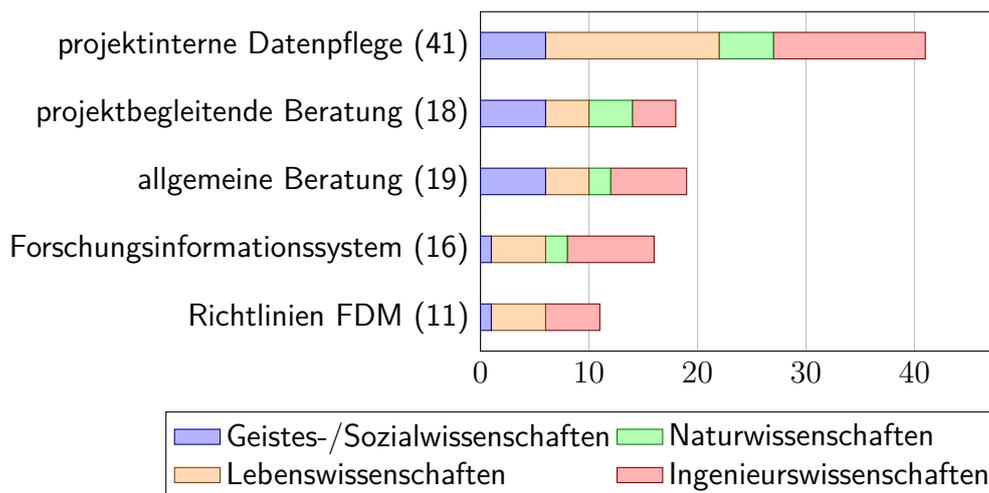


Abb. 28: Verteilung der User-Stories im Bereich Projektdokumentation

Eine strukturierte Projektverwaltung wird jedoch nicht nur für die eigenen Projekte gewünscht, sondern auch von anderen Projekten eingefordert. Das Bewusstsein, dass eine Bereitstellung projektspezifischer Informationen durch Metadaten möglich ist, ist grundsätzlich weit verbreitet. Viele Befragten erachten die vorhandenen Richtlinien zu diesem Thema jedoch als zu unspezifisch, sodass deren Umsetzung in der Praxis als schwierig angesehen wird. Daher wird vielfach gewünscht, dass Datenbanken oder Wikis zu Projekten angelegt werden, die nicht nur das erworbene Wissen innerhalb der Arbeitsgruppe zugänglich halten, sondern auch den Zugriff auf die Informationen anderer Arbeitsgruppen zulassen.

Metadatenanreicherung (84) Ein zweiter wichtiger Aspekt der Projekt- und Datendokumentation ist diese Anreicherung der Daten mit Metadaten, dem 84 unterschiedliche User-Storys zuzurechnen sind. Nur eine einheitliche und transparente Verschlagwortung kann eine standardisierte Vorhaltung und eine gute Auffindbarkeit der Informationen gewährleisten. Obwohl es in diesem Bereich einige Richtlinien zur Metadatenvergabe gibt, stellt die Umsetzung dieser Richtlinien offensichtlich eine Hürde im Forschungsalltag dar. Dabei kann der Aspekt der Metadatenanreicherung in drei größere Abschnitte unterteilt werden:

- der Wunsch nach einem standardisierten fachspezifischen Metadatenchema (24)
- personelle Unterstützung bei der Vergabe von Metadaten (17)
- Unterstützung bei der automatisierten Vergabe von Metadaten während des Entstehungsprozesses der Daten (43)

In [Abb. 29](#) erkennt man für diese drei Felder einen Trend zwischen den Wissenschaftsbereichen. Während die Geistes- und Sozialwissenschaften eher standardisierte Metadaten-Schemata als automatisierte Anreicherung benötigen, ist es bei den anderen Bereichen umgekehrt. In den einigen Fachgebieten liegt der Fokus fast ausschließlich auf der automatisierten Art der Problemlösung.

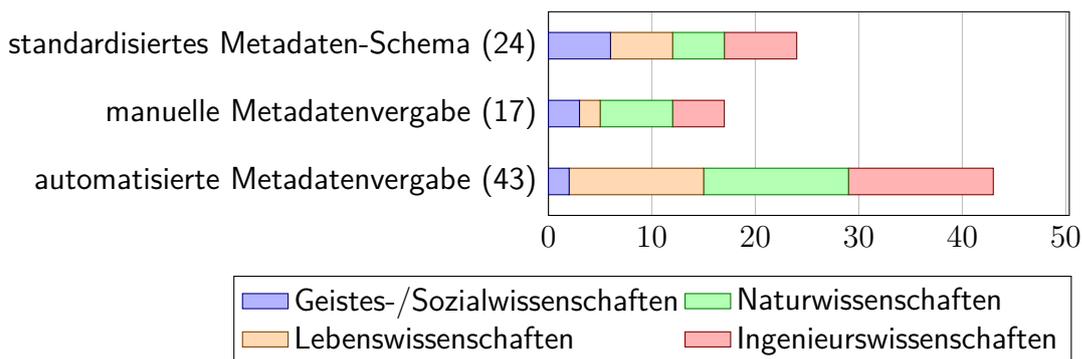


Abb. 29: Verteilung der User-Storys im Bereich Datendokumentation

Die Art und der Umfang der Metadaten, die vergeben werden sollen, hängen stark von den jeweiligen Projekten ab. Die gewünschten Metadaten werden nicht nur für die laufenden Projekte als wichtig erachtet, sondern sollen auch einen wesentlichen Schritt hin zu einer Langzeitarchivierung der Daten und deren bessere Auffindbarkeit und Nachnutzung vollziehen.

Management und Archivierungsvorbereitungen für Daten innerhalb von Projekten

Das Problem der Langzeitarchivierung, das auch im ESFK angesprochen wird, ist den Forschenden an den Universitäten zumeist sehr präsent, und es besteht die Hoffnung, durch eine begleitende Projekt- und Datenmanagementstruktur den Arbeitsaufwand für die Vorbereitung der Langzeitarchivierung zu reduzieren. Hier ist es vor allem ein – möglicherweise subjektiver – Mangel an Information und Beratung, der die Einstiegshürden in professionelle

Datenmanagementsysteme sehr hoch erscheinen lässt, zumal zeitliche und finanzielle Ressourcen investiert werden müssen, um zu Informationen und Beratung sowie letztendlich zur Integration vorhandener Modelle in der Praxis zu gelangen.

Bezugnehmend auf das Projekt- und Datenmanagement wurde in den Interviews nach vorhandenen Lösungen gefragt (s. [Abb. 30](#)).

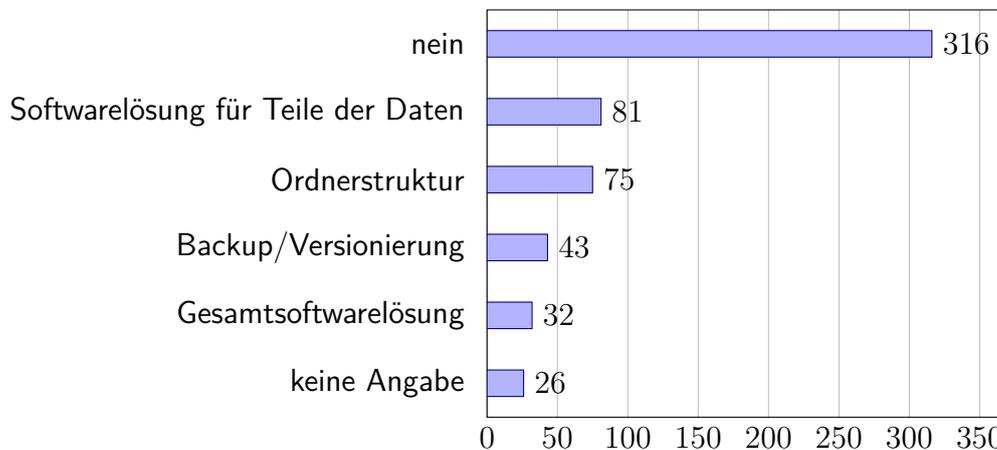


Abb. 30: „Nutzen Sie besondere Dienste oder Programme, um Ihre Daten zu verwalten?“ (Frage 17a)

Die Auswertung dieser Frage ergab, dass ein Großteil der Forschenden entweder keine konkreten Anwendungen zur Verwaltung ihrer Daten bzw. ihres Projektes verwendet oder man sich auf eine durch die Arbeitsgruppe festgelegte Ordnerstruktur geeinigt hatte.

Wenn Softwarelösungen zur Verwaltung der Daten zum Einsatz kommen, betreffen diese entweder nur Teilbereiche des Projektes bzw. der Daten oder das Backup der Daten. Häufige Beispiele für Datenhaltung von Teilbereichen sind die Vorhaltung in Bilddatenbanken oder die Verwaltung sekundärer Daten, die im Laufe des Projektes anfallen, wie etwa Literaturdatenbanken.

Der Hauptgrund für fehlende Datenmanagementstrukturen liegt in der mangelnden Notwendigkeit, ein entsprechendes Datenmanagementsystem einzuführen. Zusätzlich wird der Zeitaufwand zur einfachen Datenverwaltung als sehr hoch eingeschätzt (vgl. [Abb. 31](#)). Erklärend wurde zu dieser Angabe angemerkt, dass ein strukturiertes Projekt- und Datenmanagement weder von der entsprechenden Fachcommunity noch von den Mittelgebern honoriert wird. Eine Wertschätzung dieser Instanzen müsste sich nach Meinung der Befragten durch eine Einbeziehung des FDM in das Zitationsranking oder als positiver Faktor bei der Bewilligung von Projekten niederschlagen. Die Bemerkungen aus den Interviews geben einen Hinweis darauf, weshalb bis jetzt von einer professionellen Datenhaltung Abstand genommen wurde. Eine ausführliche Behandlung des Themas findet in [2.3. Wissenschaftliche Datenkultur](#) statt.

5.1.2.1 Forschungsinformationssystem (16)

„Als Forschende, wünschen wir uns ein universitätsinternes bzw. landesweites Forschungsinformationssystem über laufende und abgeschlossene Projekte, um über ein Webportal die

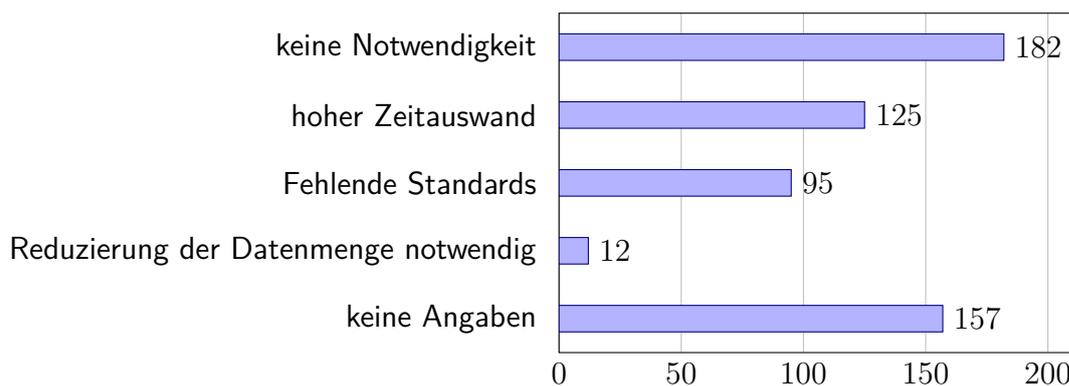


Abb. 31: Hinderungsgründe für strukturiertes Datenmanagement (Frage 17b)

Möglichkeit zu haben, andere Forschungsgruppen in unserem Fachbereich mit ähnlichen Fragestellungen einfach zu finden und kontaktieren zu können.“

Forschungsinformationssysteme sind an vielen Universitäten bereits vorhanden. Die in ihnen enthaltenen Informationen sind jedoch meist sehr heterogen und der Informationsgehalt ist stark abhängig von den zum Teil sehr unterschiedlichen Eingabeoberflächen. In der Summe steht bei diesem Wunsch die Möglichkeit im Vordergrund, die eigene Arbeitsgruppe besser und einfacher mit anderen, ähnlich arbeitenden Gruppen vernetzen zu können. Dazu ist es notwendig, detailliertere Informationen, bspw. zu methodischen Aspekten anderer Forschergruppen, auffindbar zu machen, während administrative Informationen, etwa zur Projektlaufzeit oder zum Mittelgeber, eher von untergeordneter Bedeutung für die Forschenden sind. Die formulierten Wünsche zeigen, dass solche Forschungsinformationssysteme mit entsprechender Informationstiefe teilweise bereits projektintern (4) einen Bedarf darstellen, aber universitätsintern bzw. außerhalb des Uninetzes nicht abrufbar sind (8). Nur in einem Fall wurde der Wunsch nach einer Möglichkeit zur Verwaltung und Bereitstellung von Vorlesungsdaten, sprich Lehrmaterialien, formuliert.

Neben einer Steigerung der Information über die Tätigkeiten anderer Arbeitsgruppen ist dieser Wunsch jedoch vor dem Hintergrund zu betrachten, dass Daten abgeschlossener Projekte ausgetauscht und nachgenutzt werden können (7). Dieser Aspekt ist auch für die Projektantragsphase interessant, da man mit einem besseren Überblick über laufende und abgeschlossene Projekte die Möglichkeit erhielte, selbst ein aussichtsreiches Projekt mit neuer Forschungsfrage oder –methode zu formulieren und somit Redundanzen zu vermeiden.

5.1.2.2 Richtlinien zum Projekt-FDM (11)

„Als Forschende, wünschen wir uns verbindliche Richtlinien und Anleitungen unserer Fachcommunitys bzw. der Förderinstitutionen (DFG, MWK etc.) zur Verwaltung und Archivierung unserer Forschungsdaten während und nach Projektende, um deren Zugänglichkeit und Auffindbarkeit langfristig und in einer qualitativ standardisierten Form gewährleisten zu können.“

Den inhaltlichen Kernpunkt dieser User-Stories bildet der Wunsch nach verbindlichen Richtlinien und Anleitungen wie Forschungsdaten im Laufe ihres Lebenszyklus besonders effektiv verwaltet werden sollen, da bei den Interviewten der Eindruck vorherrscht, dass es an Mög-

lichkeiten mangelt, das theoretisch wichtige FDM auch in der Praxis umzusetzen. Um für eine größere Fachcommunity zu einem verbindlichen Ergebnis zu kommen, sprich die Auffindbarkeit und Nachnutzung der Daten gewährleisten zu können, müssten nach Meinung der Befragten, die Mittelgeber, in Zusammenarbeit mit den Fachgemeinschaften, Vorgaben entwickeln.

5.1.2.3 Allgemeine Beratung zu Projekt-FDM (19)

„Als Forschende mit großem Interesse an FDM, wünschen wir uns eine zentrale Anlaufstelle die Informationen zu FDM (beispielsweise auf einer Webseite) bereitstellt, um Informationsangebote einfacher auffinden zu können und die Möglichkeit zu haben, Ansprechpartner zu fachspezifischen FDM-Fragen kontaktieren zu können.“

Die Wünsche, die unter Beratung und Information zusammengefasst sind, schließen unterschiedliche Wünsche zu dieser Thematik ein, die sich jedoch in annähernd gleichem Maße verteilen: eine Sammlung von Informationsmaterialien (8), persönliche Beratung (4) und Schulungen (6). Besonders diese Untergruppe bildet auch eine Schnittmenge mit [2.2. Informationsangebote](#). Hier wird sich allerdings auf jene Wünsche konzentriert, deren Inhalt sich ausdrücklich mit der Verwaltung von projektbezogenen Daten – im Unterschied zu Forschungs- bzw. Rohdaten – beschäftigt.

Als Einstieg wird eine Webplattform /-seite mit einer Sammlung von verschiedensten Informationsmaterialien zum Thema FDM gewünscht, da diese ansonsten als zu verstreut wahrgenommen werden. Ob diese Webseite von einer zentralen Anlaufstelle oder dem universitätseigenen Rechenzentrum bereitgestellt wird, ist dabei zweitrangig. Wichtig, und dies führt zum zweiten Punkt, ist jedoch, dass die Möglichkeit bestehen soll, auch die Kontaktdaten einer Ansprechperson für FDM über die Plattform zu bekommen. Häufig sind die Probleme, mit denen sich die Forschenden konfrontiert sehen sehr fachspezifisch, sodass eine Sammlung von Best-Practice-Beispielen zwar hilfreich ist, letztlich aber doch das persönliche Gespräch mit einem Experten bevorzugt wird. Neben dieser Sammlung von Informationen wurde auch der Wunsch nach Anleitungen in Form von Workshops und Online-Tutorials geäußert, die auf dieser zentralen Seite gesucht und abgerufen werden können.

5.1.2.4 Projektbegleitende Beratung durch FDM-Fachkraft (18)

„Als Forschende, wünschen wir uns Unterstützung durch eine FDM- oder IT-Fachkraft beim Aufbau eines Systems zur Verwaltung unserer Projekt- und Forschungsdaten, um dadurch eine Vereinfachung der administrativen Abläufe innerhalb des Projektes, aber auch die Erhaltung der Metainformationen nach Projektende zu gewährleisten.“

Eine Form der spezialisierten Beratung ist die Unterstützung durch eine FDM-Fachkraft, die als externe, jedoch dem Projekt zugewiesene, Person begleitende Aufgaben übernimmt (s. auch [3.1. Forschungsnahe Unterstützung](#)). Die mögliche Bandbreite der Tätigkeiten umfasst sowohl Beratung im administrativen Sinn, wie auch die Implementierung einer Projektmanagementsoftware und die Anreicherungen selbiger mit Daten. Die Auslagerung dieser Tätigkeiten bildet die Mehrzahl der formulierten Wünsche (15). Der Einsatz einer FDM-Fachkraft

soll dabei begleitend erfolgen und schon während der Antragsphase genutzt werden können, um die Entwicklung eines Projekt- und Datenmanagementplan zu unterstützen.

5.1.2.5 Unterstützung bei der projektinternen Datenpflege (41)

„Als Forschende, wünschen wir uns ein einfaches System zur Verwaltung unserer Forschungsdaten, um diese Daten innerhalb unserer Arbeitsgruppe strukturiert abzulegen, die Kooperation mit internationalen Partner zu vereinfachen und für die Archivierung der Daten bereits Vorarbeit zu leisten.“

Bei der projektinternen Datenpflege liegt ein deutlicher Schwerpunkt auf der Verwaltung von Forschungsdaten (17) in dem Sinn, dass eine Softwarelösung zur strukturierten Ablage der Forschungsdaten gewünscht wird. Innerhalb dieses Systems, im Idealfall einer Softwarelösung, sollen die Forschergruppen in der Lage sein, die Daten nicht nur während des Projektes aufzubewahren und auszutauschen (6), sondern die Daten auch gezielt auf eine Archivierung bzw. die Publikation (1) nach Projektende vorzubereiten. Eine Qualitätssicherung wäre an dieser Stelle ebenso wünschenswert (1). Besonders bei umfangreicheren Projekten mit einer längeren Laufzeit besteht das Problem, dass es eine systematische Verwaltung von projektspezifischem Knowhow (2) geben muss, da es anderenfalls bei Personalwechsel zu dem Verlust dieses Wissens kommt.

Für die Form, in der die Daten verwaltet werden sollen, ergaben sich aus den Interviews nur wenige Hinweise. In einigen Fällen wurde die Anlage einer Datenbank über den Datenbestand (7) angeregt, in anderen Gesprächen eine explizite Unterstützung bzw. Verbesserung bei Literaturverwaltungsmöglichkeiten (4) gewünscht.

Wie auch immer die entsprechende Softwarelösung aussieht, die initiale Dateneingabe wird oft manuell oder halbautomatisch erfolgen. Ein wichtiger Praxisaspekt dabei ist, dass auch eine projekt- bzw. institutsinterne Kontrollinstanz geschaffen werden muss, die für die gleichbleibende Qualität der Dokumentation zuständig ist. So verbinden sich diese Wünsche nach einem Verwaltungssystem, mit den Bedarfen in [3.1.2.3. Verantwortliche für Forschungsdaten](#).

5.1.2.6 Standardisiertes Metadatenchema (24)

„Als Forschende, wünschen wir uns eine nutzerfreundliche Möglichkeit (GUI) ein standardisiertes und fachspezifisches Metadatenchema vergeben zu können, um die Auffindbarkeit unserer Daten während des Projektes und nach dessen Ende zu sichern und optimale Voraussetzungen für eine potenzielle Nachnutzung zu schaffen.“

Zur Vergabe von Metadaten gibt es in verschiedenen Fachbereichen von unterschiedlichen Institutionen entwickelte Richtlinien, die die Forschenden bei der Anreicherung ihrer Daten mit Metadaten unterstützen sollen (s. [5.1.3.2. Metadaten](#)). Dennoch fehlt vielfach ein konkretes und fachspezifisches Metadatenchema, das zur Anwendung in der Praxis jedoch nötig ist (19). Abseits von Standards besteht der Wunsch für den jeweiligen Fachbereich darin, auf eine Auswahl an zu verwendenden Begriffen zurückgreifen zu können, die als Liste zur Verfügung gestellt werden soll (3). Anscheinend führt die Fülle an Möglichkeiten der Metadatenverschlagwortung zu Unklarheiten bzw. Verunsicherung beim Prozess der Vergabe der Metadaten selbst.

Neben dieser fachspezifischen Vorauswahl an Metadaten wird aber ebenfalls die mangelnde Information über die grundsätzlichen Möglichkeiten bei der Metadatenvergabe deutlich. Das äußert sich dadurch, dass Arbeitsgruppen durchaus interessiert sind auch administrative Metadaten zu ihren Projekten abzuspeichern, die Möglichkeit jedoch nicht ausschöpfen, da nicht bekannt ist, dass Metadaten schemata auch dieses Feld umfassen. Dazu ist eine umfassende Beratung zu diesem Thema erwünscht (2).

5.1.2.7 Personelle Unterstützung bei der Anreicherung mit Metadaten (17)

„Als Forschende, wünschen wir uns flexible Finanzierungsmöglichkeiten für zusätzliches Personal, das uns bei der Anreicherung unserer Daten mit Metadaten unterstützt, um den großen zeitlichen und administrativen Aufwand zu kompensieren.“

Der Wunsch nach personeller Unterstützung bei der Anreicherung der Daten mit Metainformationen formuliert vor allem, dass der zeitliche Mehraufwand, den eine solche Verschlagwortung bedeutet, kompensiert werden müsste. Da die Metadaten fach- und projektspezifisch vergeben werden müssen, um deren optimale Auffindbarkeit zu gewährleisten, muss die Vergabe meist manuell durchgeführt werden. Eine Tätigkeit, die im Moment von den Forschenden selbst geleistet wird bzw. geleistet werden müsste. Eine Entlastung in diesem Bereich kann nach Vorstellung der Befragten entweder durch die Einstellung einer weiteren Arbeitskraft (10) oder durch eine zentrale Einrichtung, die eine professionelle Archivierungsvorbereitung durchführt (7), erfolgen, die hier auch als beratende Institution auftreten kann.

5.1.2.8 Automatisierte Vergabe von Metadaten (43)

„Als Forschende, wünschen wir uns eine Softwarelösung zur automatisierten Vergabe von Metadaten, um besonders bei der Erhebung von Daten bereits fach- bzw. projektspezifische Parameter erfassen lassen zu können.“

Die automatisierte Vergabe von Metadaten, die Material- und Messparameter erfassen und vergeben, kann besonders Projekte mit einer hohen Dichte an Versuchsaufstellungen entlasten, da diese Parameter momentan entweder manuell oder gar nicht vergeben werden. Die Forschenden äußerten hierbei unterschiedliche Wünsche. Am häufigsten wurde eine Softwarelösung (19) angesprochen, um die notwendige Verschlagwortung mit Metadaten während des Projektes zu bewerkstelligen. Ein besonderes Anliegen ist dabei, dass diese automatisierte Vergabe zuverlässige und nicht manipulierbare Daten liefert, sodass eine objektive Datengrundlage geschaffen wird. Für andere Forschende (9) wäre eine Beratung bzw. Anleitung für einen optimalen Workflow jedoch ausreichend. Seltener (2) wurde die Anreicherung von Metadaten ausschließlich für Archivierungszwecke angesprochen, die erst nach Projektende erfolgt.

Die Verwendung bzw. der Wunsch nach elektronischen Laborbüchern (11) stellt eine Unterform der Forderung nach einer automatisierten Metadatenvergabe dar, da elektronische Laborbuchsystem häufig schon automatisch Parameter zu Versuchsaufbau, Messgeräten und Materialeigenschaften abspeichern. Obwohl die Verwendung dieser Systeme bereits möglich ist, scheitert es den Interviews zu Folge häufig an einer nachhaltigen Finanzierung dieser Systeme, sodass die Entwicklung von Open-Source-Lösungen in diesem Bereich bevorzugt

wird, die den Forschenden kostenfrei zur Verfügung stehen. Dennoch besteht auch Skepsis gegenüber der Datensicherheit dieser Laborbuchsysteme (1).

5.1.3 Status quo

5.1.3.1 Projektdokumentation

Aktuell ist es nicht einfach, Informationen zu einer effizienten Projektorganisation im Wissenschaftsbereich zu bekommen. Recherchiert man im Internet, so stößt man auf zahlreiche Softwarelösungen, die Projektmanagement anbieten, von denen jedoch fraglich ist, ob sie auch für wissenschaftliche Projekte geeignet sind, da sich bereits die Konzeption wissenschaftlicher Projekte von jenen in der freien Wirtschaft unterscheidet. Auch sind Lösungen zum Projektmanagement erst ab einer bestimmten Größe der Arbeitsgruppe sinnvoll einsetzbar.¹¹² Als Vorarbeiten zu einer Projektdokumentation können die Forschungsinformationsportale angesehen werden, wie sie bspw. an den Universitäten Tübingen¹¹³ und Konstanz¹¹⁴ existieren. Andere Standorte wie die Universität Freiburg¹¹⁵ oder die Universität Mannheim¹¹⁶ verfügen über Portale die allgemeinere Informationen zu den Forschungsschwerpunkten bzw. zu einzelnen Forschenden zur Verfügung stellen. Zum Teil sind solche oder ähnliche Forschungsinformationsportale noch im Aufbau. Schon jetzt ist es jedoch so, dass die Informationen, die hier abrufbar sind, sich auf administrative Faktoren beschränken. Darüber hinaus sind die Informationen, die an den jeweiligen Standorten aufgenommen werden, sehr heterogen.

Beratend fungieren an den Universitäten Einrichtungen, die besonders bei der Beantragung und Durchführung technisch orientierter Projekte unterstützten und die an manchen Standorten unter dem Schlagwort „E-Science-Center“ zusammengefasst sind.¹¹⁷

5.1.3.2 Metadaten

Richtlinien zum Aufbau und Inhalt von Metadaten werden von verschiedenen Institutionen zur Verfügung gestellt und sollen in diesem Rahmen nur exemplarisch vorgestellt werden. Die Richtlinien betreffen jedoch mitunter nur Teilbereiche der Fachgebiete und decken vor allem auch technische und administrative Metadaten ab. Die DFG hat „Praxisregeln zur Digitalisierung“ herausgegeben, die sich gezielt mit administrativen und technischen Metadaten-schemata auseinandersetzen.¹¹⁸ In der ebenfalls von der DFG herausgegebenen Denkschrift

¹¹² http://www.sgvw.ch/d/dossiers/Seiten/dossier_13_workshop_baumann.aspx [letzter Zugriff 7.5.2015]; G. Witt – M. Häcke, PMO-Implementierung in öffentlichen Forschungseinrichtungen, in: B. Sandrino-Arnt, R.L. Thomas, L. Becker (Hrsg.), Handbuch Project Management Office (2010).

¹¹³ Forschungs-Information Tübingen (FIT), <https://fit.uni-tuebingen.de/> [letzter Zugriff 7.5.2015].

¹¹⁴ SciKon, <https://scikon.uni-konstanz.de/home/> [letzter Zugriff 7.5.2015].

¹¹⁵ <http://forschdb.verwaltung.uni-freiburg.de/forschung/> [letzter Zugriff 7.5.2015].

¹¹⁶ <http://www.uni-mannheim.de/experten/> [letzter Zugriff 7.5.2015].

¹¹⁷ Bspw. an der Universität Tübingen (<http://www.escience.uni-tuebingen.de/>) oder der Universität Freiburg, angekündigt in: Anruf 1, 2015, 4-5, <https://www.rz.uni-freiburg.de/inhalt/dokumente/pdfs/anruf/anruf1501> [letzter Zugriff 15.5.2015].

¹¹⁸ http://www.dfg.de/formulare/12_151/12_151_de.pdf [letzter Zugriff 7.5.2015].

Standort	Name	Link [letzter Zugriff 17.6.2015]
Freiburg	–	–
Heidelberg	Kompetenzzentrum Forschungsdaten (KFD)	http://data.uni-heidelberg.de/
Hohenheim	Kommunikation-, Informations-, Medienzentrum (KIM)	https://kim.uni-hohenheim.de/kim-startseite
Karlsruhe	im Aufbau	–
Konstanz	Kommunikation-, Informations-, Medienzentrum (KIM)	http://www.kim.uni-konstanz.de/uebersicht/
Mannheim	–	–
Stuttgart	–	–
Tübingen	Informations-, Kommunikations- und Medienzentrum (IKM)	https://www.uni-tuebingen.de/einrichtungen/
Ulm	Kommunikations- und Informationszentrum (KIZ)	https://www.uni-ulm.de/einrichtungen/kiz.html

Tab. 5: Medien- und Informationszentren an den baden-württembergischen Universitäten

zur „Sicherung guter wissenschaftlicher Praxis“ wird der Vergabe von Metadaten kein Platz eingeräumt.¹¹⁹

Eine der bekanntesten Initiativen zur Vereinheitlichung von Metadaten ist die *Dublin Core Metadata Initiative*.¹²⁰ Dieses auf die Standardisierung bibliographischer Daten ausgelegte Metadatenformat wird zwar vielfach verwendet, ist jedoch zur Verschlagwortung von Forschungsprimärdaten kaum bzw. gar nicht geeignet. Ähnliches gilt für andere Metadatenformate wie *METS* (Metadata Encoding & Transmission Standard)¹²¹ oder *MARC 21* (Machine-Readable Cataloging).¹²²

Für Projekte mit geisteswissenschaftlichen Hintergrund hat das Projekt *DARIAH* den etwas umfassenderen Leitfaden „Fachspezifische Empfehlungen für Daten und Metadaten“ entwickelt,¹²³ der nicht nur eine Hilfestellung bei der Vergabe administrativer Metadaten geben soll, sondern sich auch mit den fachspezifischen Anforderungen an Metadaten schemata auseinandersetzt. So sind hier auch zahlreiche Links zu Organisationen zu finden, die sich in den jeweiligen Fächern mit der Vergabe von Metadaten auseinandergesetzt haben. Auch die Problematik kontrolliert-strukturierter Vokabularien wird in diesem Leitfaden thematisiert, wobei auf die Standards ISO 2788 und ISO 5964 verwiesen wird.¹²⁴ Dennoch finden sich keine konkreten Umsetzungsvorschläge.

¹¹⁹ http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf [letzter Zugriff 7.5.2015].

¹²⁰ <http://dublincore.org/> [letzter Zugriff 7.5.2015].

¹²¹ <http://www.loc.gov/standards/mets/> [letzter Zugriff 7.5.2015].

¹²² <http://www.loc.gov/marc/index.html> [letzter Zugriff 7.5.2015].

¹²³ <https://dev2.dariah.eu/wiki/pages/viewpage.action?pageId=20058160> [letzter Zugriff 7.5.2015].

¹²⁴ Diese Standards wurden mittlerweile durch die Version ISO 25964-2:2013 ersetzt: http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=53658 [letzter Zugriff 7.5.2015].

Für Daten aus sozialwissenschaftlichen Projekten bietet die Seite der GESIS¹²⁵ einen kurzen Überblick über die Thematik und empfiehlt als Metadatenschema die Data Documentation Initiative (DDI),¹²⁶ die zur Beschreibung unterschiedlicher Phasen sozialwissenschaftlicher Projekte herangezogen werden kann.

5.1.4 E-Science-Bestrebungen

Die Wichtigkeit von Forschungsdatenmanagement ist bereits im ESFK des MWK festgehalten. Hier wird ein besonderes Augenmerk auf das Datenmanagement gerichtet und hinsichtlich einer Langzeitarchivierung auf einheitliche Standards bei Dateiformaten und Metadaten gedrängt. Gerade Metadaten sind ein großes Teilgebiet, für das die User-Stories zeigen, dass eine Vereinheitlichung und Standardisierung im Sinne der Forschenden ist. Dabei kommt vor allem ein Aspekt zu tragen, der auch im ESFK angesprochen wird, nämlich eine Erweiterung bzw. Entwicklung eines Metadatenschemas, das nicht nur bibliographische und technische, sondern auch fachspezifische Attribute zur Beschreibung von Inhalt, Kontext und Provenienz der Daten enthält.¹²⁷ Was im Rahmen dieser Initiative kaum angesprochen wird, ist die Frage nach dem Management der Projekte an sich, die jedoch für die Forschenden im Allgemeinen von großen Interesse ist, und insbesondere für jene, die in einem größeren Team und / oder in internationalen Kooperationen tätig sind, da die Fülle an entstehenden Forschungsdaten innerhalb des Projektes strukturiert werden muss.

5.1.5 Gegenwärtige Handlungsempfehlungen

Projekt- und Datendokumentation sind als Themenfelder eng miteinander verbunden. Geht es bei der Datendokumentation vor allem um die kontextuelle Beschreibung der Daten, liegt der Fokus bei der Projektdokumentation vielfach auf administrativen und technischen Daten. Beides kann durch eine effizientere Vergabepaxis von Metadaten erreicht werden. Dazu sind aber Unterstützungsangebote notwendig, um die praktischen Abläufe während der zusätzlichen Informationsanreicherung zu erleichtern. Wenngleich ein breites Bewusstsein für den Mehrwert der Metadatenvergabe in der Forschungsgemeinschaft existiert, mangelt es jedoch vor allem an Zeit und Geld, diese Metadatenanreicherung in der Praxis umzusetzen.

zentrale Beratungs- und Informationswebseite Ein besonders wichtiger Punkt ist eine einheitliche Informations- und Beratungsstrategie, da sich zwar zu manchen Detailfragen Informationen finden lassen¹²⁸, es doch an einer Bündelung fehlt. Speziell für den Bereich des Projekt- und Datenmanagements könnte eine geeignete Strategie hierbei eine zentrale Beratungs- und Informationswebseite sein, die folgende Punkte abdecken könnte:

¹²⁵ <http://www.gesis.org/archive-and-data-management-training-and-information-center/forschungsdatenmanagement/dokumentation-und-metadaten/> [letzter Zugriff 7.5.2015].

¹²⁶ <http://www.ddialliance.org/> [letzter Zugriff 7.5.2015].

¹²⁷ ESFK, 81-82.

¹²⁸ Als Beispiel sei hier das Datenportal re3data.org genannt, das über verfügbare offene Daten informiert und diese zugänglich macht, <http://www.re3data.org/> [letzter Zugriff 7.5.2015].

- Sammlung von Information und Best-Practice-Beispielen (s. 2.2.5. [Gegenwärtige Handlungsempfehlungen](#))
- Planung von Schulungen und Workshops bzw. Bereitstellung von Online-Tutorien (s. 2.2.5.2. [Kurse, Workshops, Tutorien](#))
- Vermittlung von Kontaktpersonen für eine persönliche FDM-Beratung: Diese Beratung kann allgemein gehalten oder fachspezifisch sein. Eine fachspezifische Beratung könnte nach Fachbereichen getrennt auf verschiedene Standorte in Baden-Württemberg aufgeteilt werden, sodass jeder Standort die Möglichkeit hat, sich auf die Beratung bestimmter Wissenschaftsbereiche zu spezialisieren und dadurch auch qualitativ hochwertig zu unterstützen.
- Datenmanagement-Planungsinstrument (DMPI): Ein weiterer Bestandteil eines Beratungs- und Informationsportals könnte die Schaffung eines Planungsinstrumentes zum Datenmanagement, nach dem Vorbild des *Digital Curation Centre*¹²⁹ in Großbritannien sein. Ein solches Planungsinstrument ist nicht nur eine Unterstützung für die Konzeption eines Projektes, sondern auch bei der Beantragung von Forschungsmitteln, einschließlich Zusatzmitteln für FDM, die durch das DMPI leicht beantragt werden könnten.

Diese Problematik wird auch in [2.2. Informationsangebote](#) ausführlich behandelt und sollte daher hier nur kurz angesprochen werden. Über eine Beratungs- und Informationsinitiative hinaus, sind auch noch andere Ansätze zur Abdeckung des von den Befragten formulierten Bedarfs bezüglich Metadaten und Dokumentation vorstellbar.

Forschungsinformationssystem (FIS) Forschende wünschen sich primär inhaltliche Forschungsinformationssysteme (FIS), die die methodischen Aspekte der Forschung abbilden, um andere Projekte zu finden und eine Vernetzung zu fördern. Hier sollten die bereits existierenden FIS als Ausgangspunkt genommen werden. Da diese FIS sehr heterogene Informationen enthalten, ist eine landesweite Standardisierung dieser Systeme anzustreben. Im Zuge dieses Vereinheitlichungsprozesses können auch neue Inhalte implementiert werden, die dem Bedarf der Forschenden nach Forschungsinformationen entsprechen. Langfristig ist die Zusammenführung in ein landesweites FIS sinnvoll, das nicht alle administrative Interna der Universitäten umfasst, sondern eine kompatible offene Schnittmenge enthält.

Einsatz von FDM-Experten Ähnlich zu den Empfehlungen in [3.1.4. Gegenwärtige Handlungsempfehlungen](#) gibt es auch hier, bspw. für die Unterstützung zur Aufbereitung für die Publikation, Personalwünsche, die schwierig über die Einwerbung von Drittmitteln zu finanzieren sind, da es sich um punktuelle Unterstützung an verschiedenen Stellen und zu verschiedenen Zeiten im Laufe des Projektes handelt. Um diesem Problem entgegenzutreten, ist die Bündelung der Aufgaben mehrerer Projekte oder Institute bei einer FDM-Fachkraft eine mögliche Strategie. Hier sind Land und Universitäten gefordert, eine geeignete Lösung zu finden.

¹²⁹ <https://dmponline.dcc.ac.uk/> [letzter Zugriff 7.5.2015].

Workflow zur Projektdokumentation Wegen ihres Schwerpunktes auf administrativen und technischen Informationen über die Projekte, stellt der Bereich der Projektdokumentation eine wichtige und sinnvolle Ergänzung zur Metadatenanreicherung der Primärdaten dar, da diese ohne Referenzen zu Projektparametern häufig nicht nachnutzbar sind. Was jedoch fehlt, ist einerseits die Beratung zu einem entsprechenden und praktikablen Workflow und andererseits eine geeignete Softwareumgebung, die in der Lage ist, diese Daten standardisiert zu erfassen und aufzunehmen. Weiter muss eine geeignete Lösung mit entsprechenden Zugriffs- und Sicherheitsregelungen entwickelt werden, da es den Forschenden gerade während laufender Projekte besonders wichtig ist, projektspezifische Daten sicher verwahrt zu wissen.

Softwareberatung Dokumentationssoftware Besonders zur Verwaltung von Projekten und Koordination mehrerer Projektpartner gibt es bereits ein umfangreiches Softwareangebot. Hier fehlt eine Beratung, welche Lösung im Einzelfall zu wählen ist, um die hochspezialisierten Projekte an den Universitäten optimal zu unterstützen. Neben einer Beratung ist eine Empfehlung zu einer oder mehreren Softwarelösungen durch ein Fachgremium bzw. die Wissenschaftsgemeinschaft ein Weg, die Auswahl zu erleichtern. Neben einer solchen Entscheidungshilfe muss es auch Möglichkeiten einer finanziellen Unterstützung geben, die die Nutzung der empfohlenen Lösungen ermöglichen. Eine solche Unterstützung kann beispielsweise der Erwerb von landesweiten Softwarelizenzen durch die Universitäten oder das MWK sein, um den Forschenden alle notwendigen Werkzeuge zu einer professionellen Projektverwaltung zur Verfügung zu stellen.

Einbindung von Projekt- und Datendokumentationsangeboten in die Lehre Wie bereits im ESFK¹³⁰ angesprochen, sollte das Thema Datenmanagement stärker in die Lehre an den Universitäten eingebunden werden (s. auch [2.2.5.4. Forschungsdatenmanagement in der Lehre](#)). Dadurch können einerseits grundlegende Methoden und Konzepte vermittelt und andererseits die fachspezifischen Aspekte des FDM vermehrt einbezogen werden. Mittelfristig würden Forschende so einen sicheren Umgang mit dem Datenmanagement erlernen und zur Anwendung bringen. Idealerweise würden die fachlichen Lehrinhalte so in die FDM-Methodenlehre eingebettet sein, dass dort kein Abbau stattfindet.

Metadatenanreicherung Zur Anreicherung von Forschungsdaten mit Metadaten fehlen vielfach fachspezifische Richtlinien. Dieses Grundproblem kann nur durch ein verstärktes Engagement der Fachcommunitys behoben werden. Dennoch können die Fördermittelgeber durch die gezielte Förderung bzw. Einrichtung von Expertengremien einen wichtigen Impuls geben, um zu einer übergeordneten Lösung beizutragen. Ein zusätzlicher Weg wäre Repositorien zu standardisieren und zu unterstützen, sodass die Fachcommunitys sich über diese auf Standards einigen (s. auch [5.2. Repositorien für die Publikation von Forschungsdaten](#)).

Beratung Im Zusammenschluss von Bibliotheken und Rechenzentrum können vermehrt Beratung und Unterstützung zur Metadatenanreicherung angeboten und eine Infrastruktur

¹³⁰ ESFK, 88.

im Bereich Metadaten und Archivierung geschaffen bzw. Workflows und Tools für deren Umsetzung entwickelt werden. Der Aufbau von Kompetenzzentren zum Thema Metadaten kann den Prozess der Metadatenvergabe von der Projektplanung über die Datenerfassung und –aufbereitung bis hin zur Archivierung und Nachnutzung der Daten unterstützen.

Automatisierte Metadatenvergabe Ein Einstiegspunkt für eine automatisierte Metadatenvergabe können Empfehlungen der Fachcommunity für bestehende Softwarelösungen (etwa elektronische Laborbücher) sein. Durch gezielte Programme können Universitäten bzw. das MWK auch hier wieder die Umsetzung und Eingliederung in der Forschungsalltag fördern.

Personelle Unterstützung bei der Metadatenanreicherung Neben der Entwicklung geeigneter Infrastrukturen und Workflows können Kompetenzzentren auch mit einem verleihbaren Personalpool ausgestattet werden, der initial bei der manuellen Anreicherung der Daten mit Metadaten hilft, um dem Wunsch nach personeller Unterstützung dafür nachzukommen.

5.2 Repositorien für die Publikation von Forschungsdaten (243)

5.2.1 Einleitung

Ein essentieller Teil des Lebenszyklus von Forschungsdaten besteht in ihrem öffentlichen Austausch. Die Veröffentlichung von Forschungsdaten wird häufig als weiterführendes Material zu Textpublikationen, als Downloadmöglichkeit auf Webseiten, oder über strukturierte Repositorien realisiert. In einem Forschungsdatenrepositorium werden Daten gesammelt und geordnet abgelegt. Das Spektrum der gesammelten Daten hängt vom Repositorium ab. Es reicht von Institutionen, die all ihre Daten sammeln und visualisieren bis zu Fachrepositorien für Forschende einer Spezialdisziplin.

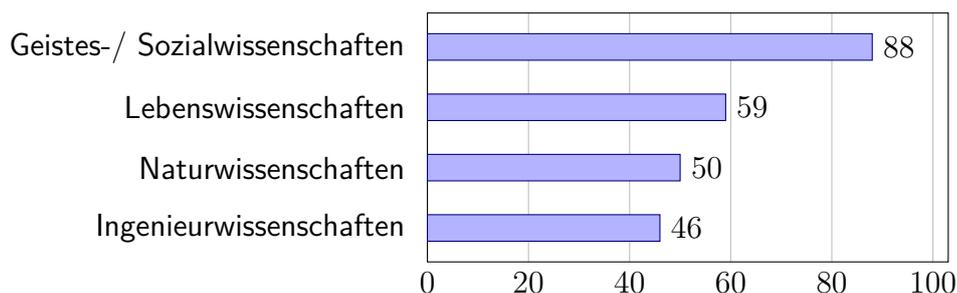


Abb. 32: Verteilung der User-Stories nach der Fachsystematik der DFG

Für Verbreitung in einer Wissenschaft spielt dabei die persistente Referenzierbarkeit von Daten eine wesentliche Rolle. Eine persistente Identifikation von Forschungsdaten ermög-

lichen beispielsweise EPIC PIDs¹³¹ oder Digital Object Identifiers (DOI) von DataCite¹³². Ein PID eines Objekts führt dabei dauerhaft und unmittelbar zum Speicherort eines Objektes. Zudem werden nachhaltige Auffindbarkeit und Nachnutzbarkeit der Daten durch eine umfassende Beschreibung mittels Metadaten möglich. Repositorien können so ihre Inhalte unveränderlich verlinken, sodass die Inhalte auch bei Umzug auf andere Stellen immernoch unter der gleichen PID zu finden sind. Einen Überblick über das stark wachsende Feld an Forschungsdaten-Repositorien bietet re3data.org¹³³.

Bei der Frage, welches Medium zur Veröffentlichung von Ergebnisdaten genutzt wird, gaben in insgesamt 569 Interviews 259 Befragte an, diese als weiterführendes Material zu textuellen Publikationen zu veröffentlichen (s. Abb. 33).

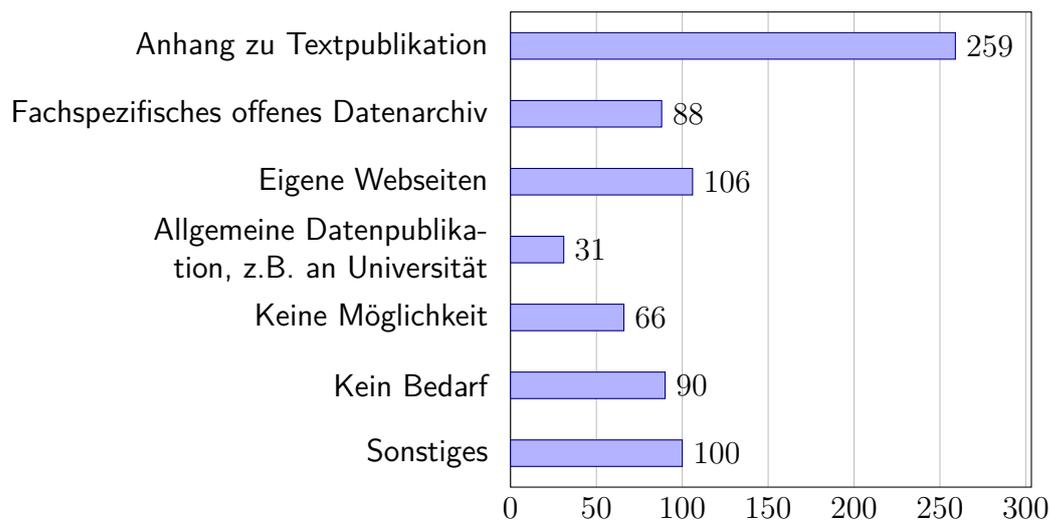


Abb. 33: „Welches Medium nutzen Sie zur Veröffentlichung Ihrer Ergebnisdaten?“ (Frage 22a – Mehrfachnennung der vorgegebenen Publikationsmöglichkeiten möglich)

106 Befragte nutzen eine eigene Webseite für die Datenveröffentlichung, 88 ein fachspezifisches offenes Datenarchiv, 31 eine allgemeine, beispielsweise universitäre Publikationsplattform. Aus 66 Interviews ging hervor, dass keine Möglichkeit für die jeweils Befragten zu einer Datenveröffentlichung bestünde. Es zeigt sich hier ein deutliches Potential für die Veröffentlichung von Forschungsdaten in strukturierten und zentral kuratierten Repositorien.

5.2.2 Bedarf der Wissenschaftsgemeinschaft

Im Folgenden werden die sich aus den Gesprächen ergebenden Aussagen zum Themengebiet der Forschungsdatenrepositorien zusammengefasst.

¹³¹ European Persistent Identifier Consortium <http://www.pidconsortium.eu>

¹³² DataCite, <https://www.datacite.org> [letzter Zugriff 11.5.2015].

¹³³ re3data.org, <http://www.re3data.org> [letzter Zugriff 11.5.2015].

5.2.2.1 Anforderungen an Repositorien (199)

„Als Forschende, wünschen wir uns Repositorien, um Daten für unsere Arbeit zu finden und anderen unsere Ergebnisse zu zeigen.“

Der Wunsch nach – vorzugsweise international – zugänglichen Repositorien lässt sich disziplinübergreifend verorten. Dabei sind neben allgemeinen Repositorien vielfach fachspezifische Repositorien erwünscht. Allerdings kommt hier hinzu, dass Forschende nicht immer an ein fachbezogenes Repository denken, sondern sich dem Thema über ihre Daten nähern. So ist ein Wunsch nach einem Repository für ganz spezifische Daten nicht ausdrücklich ein Wunsch nach einem sehr spezifischen Repository. In diesem Sinne sind die Zuordnungen zum Betriebskonzept in [Abb. 34](#) etwas schwammig. Um dieses konzeptuelle Thema dennoch zu beleuchten, haben wir uns vor allem für die Gründe der Wünsche nach spezifischen Betriebskonzepten der fachübergreifenden oder zentralen Art interessiert. *Zentrales Betriebskonzept* bedeutet hier, dass der Befragte sich explizit eine Institution (Universität, Land, Bund, DFG) als Betreiber gewünscht hat. Diese Repositorien sind oft auch fachübergreifend gemeint, meist ist das aber nicht explizit erwähnt. Ein fachübergreifendes Repository soll nach unserer Definition mindestens mehr als nur ein Fachgebiet der DFG-Systematik abdecken, wurde aber nicht mit einer zentralen Institution in Verbindung gebracht.

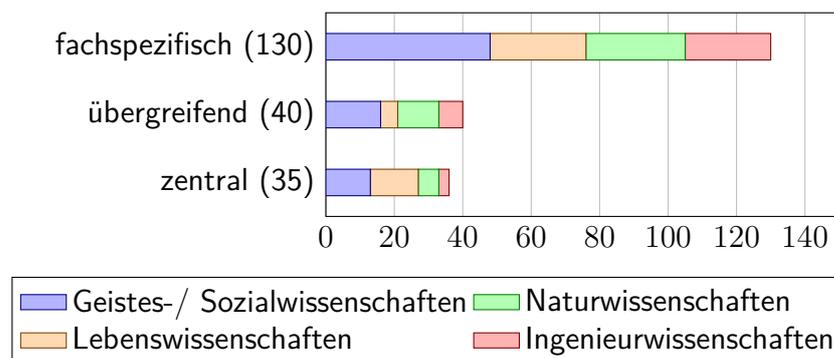


Abb. 34: Disjunkte Aufteilung nach dem Betriebskonzept des gewünschten Repositoriums in den Wissenschaftsbereichen.

In den zugrundeliegenden Daten kann man erkennen, dass es auch zwischen den Wissenschaftsbereichen Unterschiede gibt. So wünschen sich die Lebenswissenschaften überwiegend eher deshalb eine zentrale Lösung, um die Daten von vielen Journalen mit ihren Publikationen verknüpfen zu können, wohingegen in den Sozial- und Geisteswissenschaften eher der Aspekt einer erhofften Langzeitverfügbarkeit (Beständigkeit) der Daten dominiert. Die fachübergreifenden Repositorien werden von Ingenieuren und Naturwissenschaftlern hauptsächlich dazu gebraucht, um Input für ihre Modelle (Simulationsparameter wie Materialkenngrößen usw.) zu erhalten oder ihre Algorithmen an verschiedenen Problemen zu testen. Die Geistes- und Sozialwissenschaften brauchen fachübergreifende Informationen meist innerhalb ihres Wissenschaftsbereichs. Vor dem Hintergrund einer Vielzahl bereits etablierter Repositorien mit gegebenenfalls unterschiedlichen, fachspezifischen Schwerpunkten, ist eine repositorienübergreifende Suchoption bzw. ein Meta-Repository von großer Bedeutung für die Recherche

nach Forschungsdaten. Von Interviewten wurde dabei die Bedeutung einer Beschreibung von Datensätzen mittels aussagekräftiger, möglichst standardisierter Metadaten hervorgehoben.

Nach dieser Betriebskonzeptklassifikation und ihrer ersten Analyse können nun die Gründe für die jeweiligen Wünsche nach einem bestimmten Konzept analysiert werden. Diese sind in [Abb. 35](#) zusammenfassend und nach Betriebskonzepten gestapelt abgebildet.

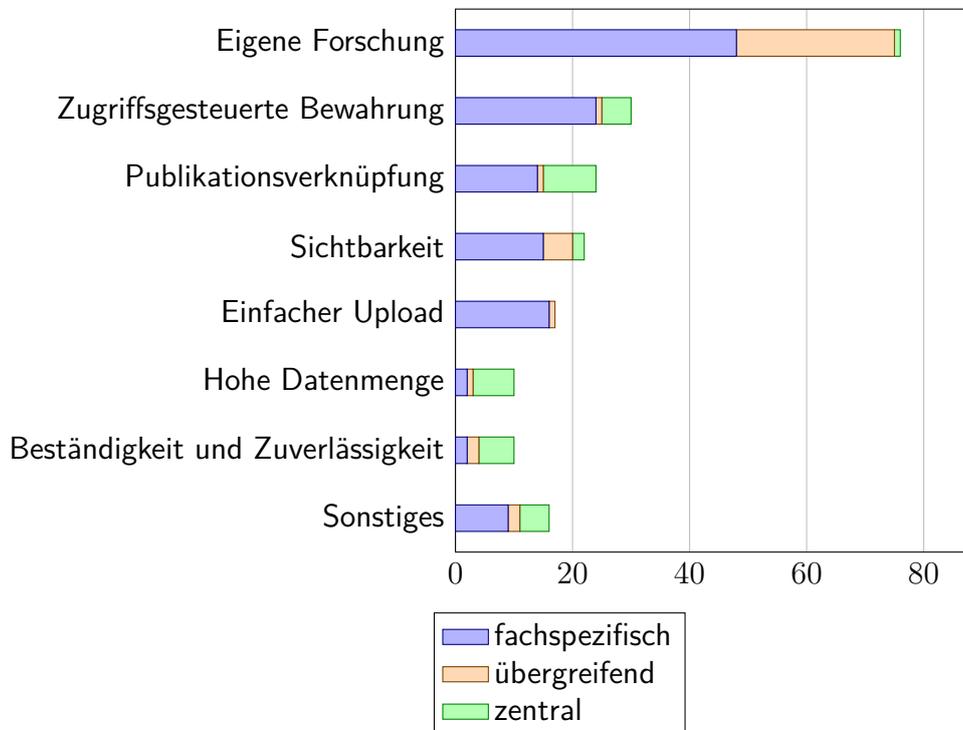


Abb. 35: Hauptzweck einer User-Story mit Repositoriumswunsch und dazugehöriges Betriebskonzept.

Es zeigt sich, dass der Hauptzweck von fachübergreifenden Repositorien darin besteht, Daten für die eigene Forschung zusammenzutragen. Die zentralen Lösungen werden dagegen kaum zur eigenen Forschung benötigt. Gründe für die Bevorzugung eines zentralen Repositoriums lassen sich nicht auf einen einzigen Aspekt reduzieren. Lediglich, dass dem zentralen Repositorium zugetraut wird publikationsrelevante Daten für alle Journale zusammen mit einem verknüpften Artikel zu publizieren, sticht ein wenig hervor. Die restlichen Hauptanforderungen, für die zentrale Lösungen gewünscht werden, sind Beständigkeit und Zuverlässigkeit sowie eine effektive Zugriffskontrolle. Auch um große Datenmengen abzulegen, scheint eine zentrale Lösung bevorzugt, wobei damit in den User-Stories oft auch ein Wunsch nach kostenfreier Ablage verknüpft war. Diese Kombination anbieten zu können, wird wohl vornehmlich der zentralen Lösung zugetraut. Weder einer zentralen Lösung noch einem fachübergreifendem Repositorium werden am ehesten eine einfache Uploadmöglichkeit zugetraut.

Besonderes Gewicht wird einem zuverlässigen Zugriffsrechtmanagement beigemessen. Dieses umfasst sowohl die Kontrolle über den Datenabruf im Zuge laufender Projekte als auch Restriktionsoptionen nach deren Abschluss, um rechtlichen Anforderungen zu genügen.

Auch der Bedarf an Repositorien für personenbezogene Forschungsdaten, die man dennoch online nutzen kann, ist groß. Oft wird auch eine gewisse Karenzzeit, in der Forschende nach Abschluss ihrer Projekte exklusives Nutzungsrecht an eigenen Daten besitzen, als wünschenswert erachtet. Ein weiterer, vielfach hervorgehobener Aspekt manifestiert sich in der Benutzerfreundlichkeit. Diese umfasst klare und übersichtliche Strukturierung zukünftiger Repositorien als auch Optionen, die die Suche nach Daten erleichtern.

5.2.2.2 Verfügbarkeit von Forschungsdaten (15)

„Als Forschende, mit Interesse an qualitativ hochwertigen Daten als Input für unsere Forschungsprojekte, wünschen wir uns einen verbesserten Zugang zu Forschungsdaten und deren persistente Referenzierbarkeit, um einen umfassenderen Überblick über den aktuellen Forschungsstand und eine breitere Datengrundlage für unsere Forschung zu haben.“

Im Rahmen wissenschaftlichen Arbeitens stellt die Verfügbarkeit von Forschungsdaten einen zentralen Faktor dar. Um sich einen umfassenden Überblick verschaffen zu können, sollten Informationen möglichst zugänglich und einfach auffindbar sein. In den Interviews wurde die Bedeutung dauerhafter Identifizierung von Forschungsdaten hervorgehoben. Es ist der Wunsch nach einer Unterstützung bei der flächendeckenden Vergabe von persistenten Identifizierungen geäußert worden.

5.2.2.3 Datenpräsentation (12)

„Als Forschende, wünschen wir uns anpassungsfähige, graphische und kontextbezogene Darstellungsmöglichkeiten von Daten in Repositorien, um die Betrachtung von Daten schon im Repositorien vorzunehmen.“

Eine Weiterentwicklung der Darstellungsmöglichkeiten von Forschungsdaten in Repositorien trägt dazu bei, die Benutzerfreundlichkeit von Repositorien zu erhöhen. In diesem Zusammenhang wird vermehrt die Bedeutsamkeit von Formatkompatibilitäten hervorgehoben. Auch die Einführung eines Standardformats für publizierte Forschungsdaten wurde diskutiert. Angesprochen wurde des Weiteren eine elektronische Hinterlegung der Daten zu Schaubildern in textuellen Publikationen.

5.2.3 Status quo

In diesem Abschnitt soll der Status quo der Forschungsdatenrepositorien in Baden-Württemberg dargestellt werden. Dazu wird zunächst re3data.org, ein repositorienübergreifende Verzeichnis, vorgestellt. Anschließend soll auf universitäre Repositorien in Baden-Württemberg eingegangen werden.

re3data.org Einen umfangreichen Überblick über Datenrepositorien liefert die Plattform re3data.org. Mit über 1000 registrierten Repositorien unterstützt re3data.org die Orientierung in der heterogenen Landschaft der Forschungsdatenrepositorien. Das durch die DFG geförderte Projekt entspringt einer Kooperation zwischen der Humboldt Universität Berlin,

dem Deutschen Geo-Forschungszentrum (GFZ) und dem Karlsruher Institut für Technologie (KIT). Seit 2014 gibt es Bestrebungen, den Dienst durch eine Zusammenarbeit mit der Purdue University (USA) auszubauen.

Des Weiteren wurden Initiativen ergriffen, institutionelle fachübergreifende Repositorien aufzubauen, wie die Angebote baden-Württembergischer Universitäten, FreiDok plus, HeiDATA und MADATA. Über das Horizon 2020 Projekt EUDAT steht auch ein europaweites Forschungsdaten-Repositorium zur Verfügung. Im Folgenden sollen Charakteristika dieser Plattformen beleuchtet werden.

Universitäre Repositorien in Baden-Württemberg Folgend wird auf diejenigen Datenrepositorien an Baden-Württembergischen Universitäten näher eingegangen, die eine disziplinunabhängige Datenpublikationsmöglichkeit darstellen (Stand Juni 2015). Der aktuelle Ausbau des Forschungsinformationssystems der Universität Freiburg, FreiDOK plus¹³⁴, sieht u.a. das integrierte Angebot eines Repositoriums zur Veröffentlichung von Forschungsdaten vor. Das System gewährleistet die Verknüpfung von Informationen zu Publikationen aller Art, zu Personen und Institutionen. Das an der Universität Heidelberg initiierte und vom dortigen Kompetenzzentrum Forschungsdaten kuratierte Repositorium HeiDATA¹³⁵ ermöglicht die Publikation und Archivierung von Forschungsdaten. Persistenter Zugriff und Zitierbarkeit werden mittels DOI-Kennungen realisiert. Das Repositorium ist zudem in den Data Citation Index integriert. MADATA¹³⁶ ist ein von der Universität Mannheim eingerichtetes Forschungsdatenrepositorium. Der kostenlose Dienst richtet sich fakultätsübergreifend an alle Forschenden der Universität Mannheim und ruft diese dazu auf, ihre Forschungsdaten auf der Plattform zugänglich zu machen. Dabei können die Benutzenden individuell über Zugangsmöglichkeiten und -beschränkungen entscheiden. Dokumente werden mit einer DOI-Kennung versehen.

Horizon 2020 Repositorium Forschende aus Europa können B2SHARE¹³⁷ von EUDAT zum öffentlichen Austausch kleinerer Datenmengen frei verwenden. Für größere Datenmengen wurden entsprechende Kostenmodelle in Form von Premiumangeboten entwickelt. Die Plattform bietet konfigurierbare Zugänglichkeit der Daten mit persistenten Identifizierung, fachspezifische Metadaten sowie REST APIs für die Repräsentation auf anderen Internetseiten. Zudem wird langfristige und sichere Aufbewahrung der Daten gewährleistet.

ZENODO Das CERN stellt seine Speicherkapazitäten teilweise für das allgemeine Datenrepositorium ZENODO¹³⁸ zur Verfügung. Es ist als Auffang-Repositorium gedacht, falls sonst kein geeignetes fachspezifisches Repositorium zur Verfügung steht, weshalb der Service kostenlos für kleine Wissenschaftsdisziplinen ist. Die Publikationslizenz lässt sich frei wählen, wobei offene Lizenzen bevorzugt behandelt werden. Die Maximale Datensatzgröße ist zunächst 2GB, lässt sich aber gegebenenfalls erweitern.

¹³⁴ FreiDOK, <http://www.freidok.uni-freiburg.de> [letzter Zugriff 11.5.2015].

¹³⁵ HeiDATA, <https://heidata.uni-heidelberg.de> [letzter Zugriff 11.05.2015].

¹³⁶ MADATA, <http://madata.bib.uni-mannheim.de> [letzter Zugriff 11.05.2015].

¹³⁷ B2SHARE, <http://eudat.eu/> [letzter Zugriff 11.05.2015].

¹³⁸ <http://zenodo.org/> - [letzter Zugriff 30.6.2015].

5.2.4 E-Science-Bestrebungen

Im ESFK wird insbesondere die Bedeutsamkeit von Repositorien hervorgehoben:

„Bei der Entwicklung effektiver Arbeitsumgebungen schafft sich die Wissenschaft eine integrierte digitale Plattform, in der alle Daten, alle Literatur, alle Berechnungsmethoden und weitere Ressourcen verfügbar sind.“ (ESFK, 8)

Um jene infrastrukturellen Rahmenbedingungen weiter auszubauen, wird die Einrichtung eines landesweiten Forschungsdaten-Repositoriums vorgeschlagen. Dieses soll über ein umfassendes Metadatenschema verfügen, Optionen der Archivierung von Daten bereitstellen und eine persistente Adressierung von Dokumenten ermöglichen. Darüber hinaus werden Datenschutz und Qualitätssicherung höchste Priorität beigemessen. Letztere soll u. a. durch ein Zertifikat für die Nachvollziehbarkeit und Reproduzierbarkeit von Forschungsdaten gesichert werden. Damit ein landesweites Forschungsdatenmanagement realisiert werden kann, sieht das MWK eine Kooperation zwischen Infrastruktureinrichtungen und Forschenden als unerlässlich an. Im Konsens mit den Empfehlungen der DFG¹³⁹, die zentralen Speichermöglichkeiten eine deutliche Überlegenheit gegenüber lokalen Lösungen bescheinigen, lässt sich die Bedeutsamkeit von Repositorien unterstreichen.

5.2.5 Gegenwärtige Handlungsempfehlungen

Im Vergleich zur Bedarfserhebung lässt sich feststellen, dass die etablierten Repositorien den benannten allgemeinen Grundanforderungen weitestgehend genügen. Für einige Forschende ist der Bedarf entweder an überregionalen bzw. institutionellen Repositorien bislang nicht abgedeckt oder die entsprechenden Angebote sind nicht bekannt.

Bekanntmachung von Repositorien Um für den zweiten Fall Repositorien weiter zu etablieren, ließe sich die Sichtbarkeit durch Verzeichnung in diversen Indizes, wie re3data.org oder auch den Data Citation Index¹⁴⁰ der Plattform Web of Science¹⁴¹ erhöhen. Eine weitere Vernetzung mit Textpublikationsorganen oder übergeordneten Plattformen einer jeweiligen Community anzustreben, würde ebenso förderlich wirken.

Ausbau von Repositorien In Fällen des Mangels eines passenden Repositoriums für Forschungsdaten einzelner Communitys entsteht ein Impuls aus den einzelnen Communitys heraus zur Etablierung neuer, spezifisch angepasster Datenpublikationsplattformen, die die Daten in ihrem jeweiligen Forschungskontext präsentieren können. Für die Behandlung speicherintensiver Datensätze, die an zentraler Stelle effizient gehalten werden können, ist dabei ein transparentes Betriebs- und Kostenmodell gefragt, wenn die zentralen Infrastruktureinrichtungen diese Repositorien unterstützen wollen.

¹³⁹ Deutsche Forschungsgemeinschaft: Informationsverarbeitung an Hochschulen – Organisation, Dienste und Systeme, Empfehlungen der Kommission für IT-Infrastruktur für 2011-2015. Bonn, 2010.

¹⁴⁰ Data Citation Index, http://wokinfo.com/products_tools/multidisciplinary/dci/ [letzter Zugriff 11.5.2015].

¹⁴¹ Web of Science, <http://wokinfo.com> [letzter Zugriff 11.5.2015].

Repositorienübergreifende Suche nach Forschungsdaten Entsprechend der von Forschenden geäußerten Wünsche nach einer übergreifenden Durchsuchbarkeit von Repositorien, können Entwicklungen zur Verfügungstellung einer Art Metasuchmaschine für Forschungsdaten initiiert werden. Als Grundvoraussetzung kann hier ein einheitliches Metadatenschema der damit abgedeckten Repositorien genannt werden oder zumindest eine standardisierte Schnittstelle der Repositorien für Suchanfragen.

5.3 Archivierung (108)

5.3.1 Einleitung

Der Begriff der digitalen Langzeitarchivierung (in Kürze auch nur Archivierung genannt), wird teilweise unterschiedlich interpretiert. Von Forschenden wird nicht selten darunter die langfristige Speicherung entweder auf internen / externen Festplatten, CDs / DVDs oder früher auch auf Disketten verstanden. Dies stellt jedoch keine geeignete langfristige Archivierung dar. Archivierung umfasst idealerweise auch die Absicherung der Auffindbarkeit und Interpretation der Daten trotz technischen und sonstigen Veränderungen. Den Forschenden ist allerdings die Notwendigkeit der Archivierung – ungeachtet der korrekten Umsetzung – aus verschiedenen Gründen wichtig, weshalb sich das Thema in vielen Fachbereichen präsent ist. Wie der [Abb. 36](#) entnommen werden kann, stellen nicht die historischen Fächer der Geistes- und Sozialwissenschaften, sondern die Lebenswissenschaften den signifikant größten Block der vorhandenen User-Stories. Es lässt sich an unseren Daten ablesen, dass auch die neueren Fächer rund um die Bioinformatik die Notwendigkeit der Archivierung sehen, jedoch oft auch aufgrund ihrer Datenmengen mehr Probleme bei der Archivierung haben.

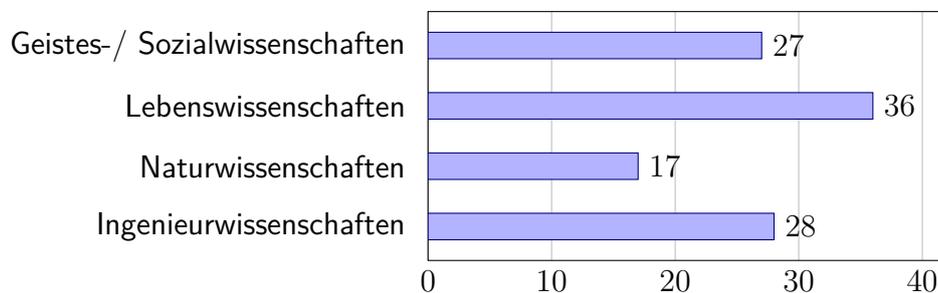


Abb. 36: Verteilung der User-Stories nach der Fachsystematik der DFG

Wie anhand der Bedürfnisse in [5.3.2.1. Allgemeine Langzeitarchivierung](#) zu sehen ist, decken sich die Wünsche zur Archivierung mit den DFG-Vorgaben. Der vorgesehene Zeithorizont stellt jedoch keine harte Grenze dar. Der Aspekt der Metadatenanreicherung ist von zentraler Bedeutung, da ohne geeignete Metadaten die archivierten Daten nicht geeignet gefunden und nachgenutzt werden können. Diese Arbeiten müssen im Zusammenhang mit der Vorbereitung zur Langzeitarchivierung von Projekt- und Forschungsdaten gesehen werden und werden daher in [5.1. Projekt- und Datendokumentation](#) behandelt. Aus diesem Grund wird in diesem Kapitel auf Metadaten nicht näher eingegangen. Im Folgenden werden die Wünsche, Probleme und Anregungen der Befragten analysiert.

5.3.2 Bedarf der Wissenschaftsgemeinschaft

Die 108 zugrundeliegenden User-Stories untergliedern sich in sieben Themen. In [5.3.2.1. Allgemeine Langzeitarchivierung](#) geht es um Wünsche zur generellen Archivierung. Die genaue Archivierungsform ist hierbei den Forschenden nicht wichtig. Oftmals wird sich jedoch auch die Archivierung der Daten innerhalb eines Repositoriums gewünscht. Dieser Aspekt nach einer konkreten Archivierungsform wird separat in [5.3.2.2. Langzeitarchivierung in Repositorien](#) erfasst. Ein weiteres Thema betrifft die langfristige Zugriffssicherung ([5.3.2.3. Langfristige Zugriffssicherung](#)), bei der es vor allem darum geht, die Forschungsdaten auch noch zu einem späteren Zeitpunkt verwenden zu können, obwohl die verwendete Software oder die Formate über die Zeit obsolet geworden sind. Die Forschenden wünschen sich jedoch auch, dass die Tätigkeit der Archivierung von einer anderen Stelle für sie übernommen wird. Dieser Aspekt wird in [5.3.2.4. Outsourcing](#) erläutert. Um eine eigene Archivierung durchführen zu können, wünschen sich die Forschenden sowohl genügend finanzielle Mittel ([5.3.2.5. Finanzierung](#)) als auch geeignete Werkzeuge ([5.3.2.6. Hilfstoos](#)). In [5.3.2.7. Allgemeine Diskussion der Langzeitarchivierung](#) geht es um grundlegende Gedanken, die einige Forschende zu diesem Thema haben.

Um den allgemeinen Bedarf der Archivierung zu erfassen, wurden im Interviewleitfaden zwei Fragen (19a und 19b, [Abb. 37](#) und [Abb. 38](#)) zur Langzeitarchivierung gestellt. Zur Veranschaulichung wurde in [Abb. 37](#) neben den möglichen Optionen zusätzlich „*Nein*“ hinzugefügt, falls der erste Teil der Frage – ob die Langzeitarchivierung sinnvoll wäre – mit Nein beantwortet wurde. Des Weiteren wurde in der Abbildung „*Keine Aussage möglich*“, „*Keine Antwort*“ und „*Ja (ohne weitere Nennung des Zweckes)*“ hinzugefügt. „*Keine Aussage möglich*“ bedeutet, dass der Interviewpartner sich nicht sicher war und aus diesem Grund keine definitive Einschätzung abgeben konnte, wohingegen „*Keine Antwort*“ bedeutet, dass die Frage dementsprechend gar nicht beantwortet wurde.

Wie der Verteilung der Antworten entnommen werden kann, ist zu sehen, dass 95 Interviewte (entspricht 16,66 %) eine Langzeitarchivierung als nicht sinnvoll erachten, 28 (4,91 %) keine Antwort gaben und 9 (1,58 %) keine Aussage zur Langzeitarchivierung treffen konnten. Somit finden mehr als dreiviertel der Interviewten die Langzeitarchivierung für sinnvoll¹⁴², wobei die Archivierung als Informationsquelle für zukünftige Forschung, zur Beweisbarkeit der Resultate, für den späteren Eigenbedarf sowie zum Vergleich mit ähnlichen Daten als ähnlich wichtig empfunden wird. Am häufigsten wird die Archivierung als Informationsquelle für zukünftige Forschung angesehen. Der Unterschied wird deutlicher, sofern der prozentuale Anteil auf die Anzahl der Antworten (bezeichnet als $\text{Prozent}_{\text{Antworten}}$), als auch auf die Gesamtzahl der Interviews (bezeichnet als $\text{Prozent}_{\text{Interviews}}$) betrachtet wird (s. [Tab. 6](#)). Insgesamt wurden demnach pro Interview im Schnitt 1,7 Antworten gegeben, wobei ersichtlich ist, dass nahezu die Hälfte der Befragten die Archivierung als Informationsquelle für zukünftige Forschung sehen. Es gilt jedoch zu beachten, dass die Antworten unabhängig von der konkreten Archivierungsform sind. Eine notgedrungene Schwäche der Frage ist, dass zwar zehn Jahre erwähnt werden, aber die Befragten vielleicht dennoch eher in kürzeren Zeiträumen denken. Zum Vergleich dafür dient die Nachfrage (siehe [Abb. 38](#)).

¹⁴² Dies wurde angenommen, sofern ein Kreuz bei einer der Wahlmöglichkeiten gesetzt wurde.

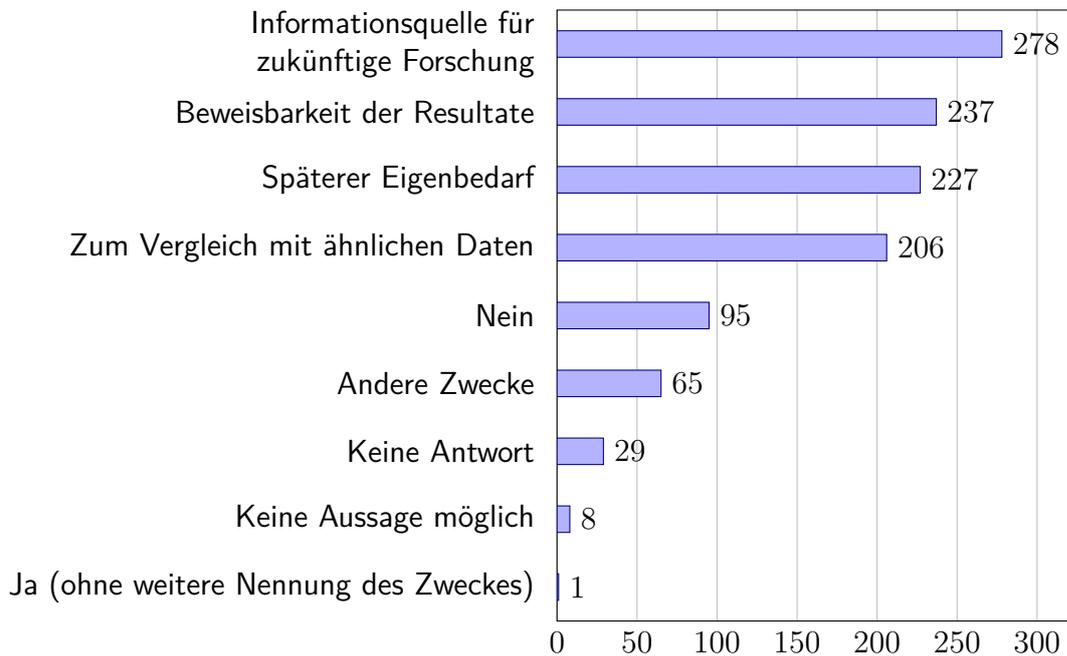


Abb. 37: „Würden Sie eine Langzeitarchivierung (> 10 Jahre) eines Teils Ihrer Daten für Sinnvoll halten? Wenn ja, wofür wäre es Ihnen besonders wichtig?“ (Frage 19a)

Antworten	Anzahl	Prozent _{Interviews}	Prozent _{Antworten}
Informationsquelle für zukünftige Forschung	278	48,77%	27,44%
Beweisbarkeit der Resultate	237	41,58%	23,40%
Späterer Eigenbedarf	227	39,82%	22,41%
Zum Vergleich mit ähnlichen Daten	206	36,14%	20,34%
Andere Zwecke	65	11,40%	6,42%
Summe	1013	177,71%	100,01%

Tab. 6: Prozentueller Anteil der Antworten bezogen auf die Anzahl der Interviews (570) und aller Antworten (1013)

Die Verteilung der gegebenen Antworten bei 570 Interviews, kann der [Abb. 38](#) entnommen werden.

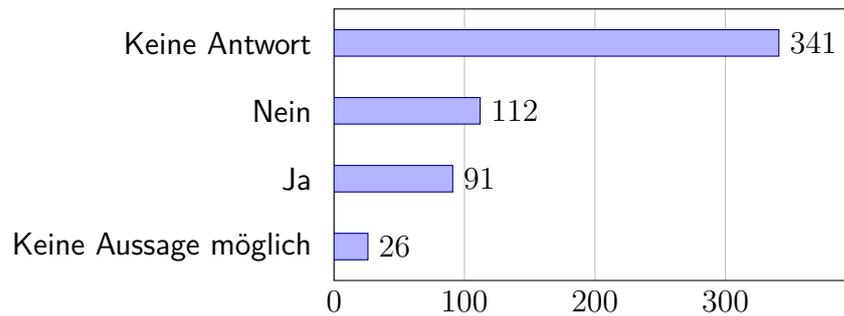


Abb. 38: „Falls "Messungen" (auch Umfragen) an "realen Dingen" oder einmalige Beobachtungen vorgenommen werden: Könnte das für jemand in 100 Jahren interessant sein? Speichern Sie diese Daten noch einmal extra gesichert ab?“ (Frage 19b)

Ob die Daten in 100 Jahren interessant sein könnten, haben aufgrund der Vorbedingungen insgesamt 341 Interviewte (59,72 %) nicht beantwortet. Weitere 26 (4,55 %) Interviewte konnten dies nicht abschätzen. Etwas mehr als die Hälfte der Übrigen (55 %) halten eine derart lange Archivierung für nicht sinnvoll, wobei der Haupteinwand war, dass bedingt durch den technischen Fortschritt die Daten irgendwann irrelevant werden und somit nur eine kürzere Archivierung Sinn machen würde. Von den 91 Interviewten, die eine Archivierung für länger als 100 Jahre als sinnvoll erachteten, haben 19 (20,88 %) den zweiten Teil der Frage, ob sie diese Daten noch einmal extra gesichert abspeichern, mit „Ja“ beantwortet.

5.3.2.1 Allgemeine Langzeitarchivierung (56)

„Als Forschende, wünschen wir uns ein Speicherarchiv, das einfach zu bedienen und ausfallsicher ist, und eine von mir bestimmbare Zugriffsbeschränkung bietet, um unsere Forschungsdaten zur Nachnutzung sichern zu können.“

In den insgesamt 56 User-Stories zum Thema der allgemeinen Langzeitarchivierung haben die Forschenden naturgemäß unterschiedliche Vorstellungen, welche Eigenschaften ein Archiv besitzen und was dieses für sie leisten soll.

Mindestens 12 Forschende legen vor allem darauf Wert, dass der Dienst sicher (6), Bit-sicher (1), ausfallsicher (2), verlässlich (2) und stabil (1) ist. Die sonstige Nichterwähnung impliziert jedoch nicht, dass diese Anforderungen für andere nicht notwendig wären, sondern vielmehr, dass dies implizit von einem Archiv erwartet wird. Oft findet eine solch explizite Erwähnung statt, wenn es zuvor ein entsprechend gelagertes Problem gab, dass dem Befragten noch präsent war.

Insgesamt 14 Forschenden ist vor allem eine Zugriffsbeschränkung auf ihre archivierten Daten wichtig. Hierbei geht es darum, selbstständig bestimmen zu können, wer Zugriff auf die Daten erhalten soll, um beispielsweise Kooperationspartnern ebenfalls einen Zugang gewähren zu können.

Insgesamt 33 Forschende haben Aussagen über die Speicherdauer getroffen. So wünschen sich sechs von ihnen, dass ihre Daten für mindestens 10 Jahre nach DFG-Anforderung ar-

chiviert werden sollen. Drei zielten auf 10 bis 15 Jahre ab, sechs weitere auf mehr als 10 Jahre. Einzelne wollten 20, 50 oder gar 100 Jahre übertreffen. 15 Forschende haben die Archivierungsdauer nicht genau spezifiziert, wünschten sich aber eine Langzeitarchivierung ihrer Daten.

Einzelaspekte für eine zentrale Lösung waren u. a. die Vertrauenswürdigkeit gegenüber Kooperationspartnern sicherzustellen, Drittmittelgeber die ordnungsgemäße Archivierung signalisieren zu können, nicht eigenständig ein Archivserver administrieren zu müssen und die Vermeidung der gängigen Praxis externe Festplatten / DVDs als Archiv zu verwenden.

5.3.2.2 Langzeitarchivierung in Repositorien (15)

„Als Forschende, wünschen wir uns ein Repitorium, das PIDs zum Zitieren der Forschungsdaten bereitstellt, um Forschungsdaten und Software langfristig und offen zur Verfügung stellen zu können.“

In insgesamt 15 User-Stories wünschen sich die Forschenden Repositorien, um ihre Forschungsdaten langzeitarchivieren und gleichzeitig anderen zur Verfügung stellen zu können. Diese Repositorien sollen idealerweise EU-Standards bezüglich Format und notwendigen Metadaten setzen und von der DFG und den Zeitschriftenverlagen akzeptiert werden. Zusätzlich ist mindestens sieben Forschenden vor allem die Zitierfähigkeit mittels PIDs der Forschungsdaten innerhalb des Repositoriums wichtig. Somit sollte das Repitorium offen sein, um Forschungsdaten auch austauschen zu können. Einzelne Wünsche waren, dass die Forschungsdaten fälschungssicher hinterlegt werden können, dass das Repitorium fachspezifisch (2) sein sollte und eine Zugriffsprotokollierung bieten sollte, um nachvollziehen zu können, wer die Daten verwendet. Der Hauptgrund für die Verwendung eines Repositoriums für die Langzeitarchivierung stellt vor allem die Nachnutzung (5) dar. Insgesamt ergibt sich die Tendenz, dass Repositorien vermehrt als Langzeitarchivierungsspeicher angesehen werden. Dieser Trend wird sich vermutlich mit einer weiteren Durchdringung von Repositorien verstärken.

5.3.2.3 Langfristige Zugriffssicherung (10)

„Als Forschende mit einer spezifischen IT-Umgebung und davon abhängigen Abläufen, wünschen wir uns Services und Rahmenbedingungen, die uns erlauben unsere Software und IT-Umgebung zu archivieren, um wirklich originalgetreu zu archivieren und einmal funktionierende Abläufe lange Zeit erhalten zu können.“

Das Ziel einer langfristigen Zugriffssicherung ist, digitale Objekte nachhaltig zugreif- und vor allem nutzbar zu halten. Falls die notwendige Software und ein entsprechend geeignetes Betriebssystem archiviert werden, kann die Funktionalität eines digitalen Objektes erhalten werden, wodurch obsoleete Formate in ihrer ursprünglichen Umgebung weiterhin verwendet werden können. Bei der üblichen Archivierung wird meistens lediglich das digitale Objekt archiviert. Ob dieses Objekt jedoch zu einem späteren Zeitpunkt geeignet verwendet werden kann, kann ohne die Umgebungserhaltung oft nicht garantiert werden.

Ein weiteres Themengebiet betrifft die Archivierung von Softwares und Formaten. Es wird sich gewünscht, dass Formate langlebiger werden sollen, um Daten, ohne Formate umwandeln zu müssen, langfristig nachnutzen zu können. Außerdem sollen nicht nur die Daten

sondern auch die verwendete Software in ihrer exakten Version mit der originalen Umgebung archivierbar sein, um die Ergebnisse mit der korrekten Softwareversion innerhalb ihrer ursprünglichen Umgebung reproduzieren zu können. Unterschiedliche Softwareversionen bzw. die Verwendung von unterschiedlichen Umgebungen kann sonst zu unterschiedlichen Ergebnissen führen.¹⁴³

5.3.2.4 Outsourcing (7)

„Als Forschende, wünschen wir uns eine zentrale Dienstleistung, die eine vollständige Übernahme der Archivierung unserer Forschungsdaten übernimmt und uns vor dem Löschen der Daten kontaktiert, um unsere Forschungsdaten über einen langen Zeitraum (> 10 Jahre) mit minimalem Aufwand archivieren zu können.“

Forschende wünschen sich hier, dass sie ihre Daten an eine (zentrale) Stelle abgeben können, die sich anschließend um die korrekte Archivierung der Daten kümmert. Somit wollen die Forschenden minimalen Aufwand investieren, die Daten jedoch langzeitarchiviert haben. Mögliche Gründe hierfür sind u. a., dass sie entweder nicht genügend Zeit und Ressourcen aufwenden können oder ihnen schlichtweg das notwendige Knowhow fehlt, die Daten entsprechend aufzubereiten. Jedoch zeigt dieser Wunsch nach einer zentralen Dienstleistung, dass den Forschenden die Wichtigkeit der Archivierung durchaus bewusst ist. Gleichzeitig sollen die Mitarbeiter dieses Services die langfristige Datenverfügbarkeit sicherstellen, vor dem Löschen der Daten den entsprechenden Forschenden kontaktieren um zu klären, ob die Daten gelöscht werden können oder noch weiter archiviert werden sollen.

5.3.2.5 Finanzierung (8)

„Als Forschende, wünschen wir uns ein standardisiertes Konzept für eine nachhaltige Finanzierungslösung, um Forschungsdaten, entgegen unserer begrenzten Mittel, langfristig (auch nach Projektende) archivieren zu können.“

Die Gründe für die Wünsche nach einer nachhaltigen Finanzierungslösung lassen sich in zwei Gruppen „generelle Archivierung“ und „Archivierung nach Projektende“ einteilen. Bei der ersten Gruppe geht es vor allem darum, größere Datenmengen von bis zu 50 TB generell langfristig, auch ohne konkreten Forschungszweck, archivieren zu können. Die Daten sollen vor allem zum Zwecke der Nachnutzung oder auch zum Forschen mit neuer Zielsetzung archiviert werden. Bei der zweiten Gruppe geht es darum, dass die vorhandenen finanziellen Mittel nur für einen gewissen Zeitraum, meistens während der Projektlaufzeit, vorhanden sind. Daher wird sich eine nachhaltige Finanzierungslösung bzw. ein standardisiertes Konzept zur Kostenübernahme für die Archivierung von Forschungsdaten auch nach Projektende gewünscht.

¹⁴³ Gronenschild, Ed HBM, et al. „The effects of FreeSurfer version, workstation type, and Macintosh operating system version on anatomical volume and cortical thickness measurements.“ PLoS One 7.6 (2012)

5.3.2.6 Hilfstools (7)

„Als Forschende, die gern ihre Forschungsdaten archivieren möchten, wünschen wir uns ein einfaches, sicheres, schnelles und zuverlässiges Werkzeug, das auch die Möglichkeit bietet, Metadaten zu den Daten hinzuzufügen, um unsere Daten (automatisch; nach Projektende) archivieren zu können.“

Generell soll ein Archivierungswerkzeug eine benutzerfreundliche, einfach zu handhabende Oberfläche (GUI) bieten und die Daten im Archiv lösungssicher sein.

Abseits der Wünsche nach Archivierungswerkzeugen wurde sich gewünscht, dass bwSync&Share eine Archivierungsfunktion bereitstellt, um seine Daten direkt über bwSync&Share einfach und mit der Übernahme der Zugriffsbeschränkungen mit einem Klick langzeitarchiviert werden können. Somit soll jeder, der vorher auf die Daten Zugriff hatte auch weiterhin auf die Daten zugreifen können. bwSync&Share bietet zwar eine Historie von Daten an, dies ist jedoch nicht mit einem Archiv zu vergleichen, da die Daten nach kurzer Vorhaltdauer wieder gelöscht werden.

5.3.2.7 Allgemeine Diskussion der Langzeitarchivierung (6)

„Als Forschende, die eine Diskussion zum Thema der Langzeitarchivierung anregen möchten, wünschen wir uns, dass die allgemeinen Archivierungsrichtlinien besser auf die verschiedenen Wissenschaftsbereiche abgestimmt werden, da eine mehrfache Archivierung bzw. eine Archivierung der Forschungsdaten zum Zwecke der Nachnutzung nicht für alle Wissenschaftsbereiche gleich sinnvoll ist.“

Die Gründe für eine allgemeine Diskussion zum Thema der Langzeitarchivierung sind naturgemäß divers. Ein Argument gegen die eigene Durchführung der Langzeitarchivierung betrifft den Erhalt von externen Daten. Da diese Daten von einem externen Partner geliefert werden und jeder verpflichtet wird, seine eigenen Forschungsdaten zu archivieren, würden die Daten offensichtlich bei dem externen Partner archiviert werden. Daher würden die erhaltenen Forschungsdaten somit mehrfach archiviert werden, was eine Verschwendung von Zeit-, Personal- und Geldressourcen darstellen würde.

Ein anderer Punkt ist, dass eine allgemein festgeschriebene Archivierungsdauer von Forschungsdaten für alle Wissenschaftsbereiche als wenig sinnvoll erachtet wird. Zum einen ergeben sich aufgrund von neuen Methodiken beispielsweise bei besseren Messmethoden höhere Messgenauigkeiten, weshalb alte Daten nicht sinnvoll nachgenutzt werden würden, zum anderen wird eine Langzeitarchivierung (> 10 Jahre) zum Zwecke der Nachnutzung teilweise als uninteressant erachtet, sobald die wissenschaftlichen Erkenntnisse sich gefestigt haben und anerkannt sind.

5.3.3 Status quo

Die Archivierung von Forschungsdaten zum Zwecke der Nachvollziehbarkeit der wissenschaftliche Ergebnisse wird schon seit Längerem von verschiedenen Stellen gefordert. So fordert die EU in ihrem Horizon 2020 Data Pilot die Archivierung der Forschungsdaten:

„A further new element in Horizon 2020 is the use of Data Management Plans (DMPs) detailing what data the project will generate, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved.“ (Guidelines on Data Management in Horizon 2020¹⁴⁴)

Diese Forderung ist jedoch nur für von der EU geförderte Data-Pilot-Projekte bindend. Allerdings werden zugleich von der DFG allgemeine Anforderungen zur guten wissenschaftlichen Praxis formuliert, die für alle Forscher gleichermaßen gelten sollen. Laut der „*Empfehlung 7: Sicherung und Aufbewahrung von Primärdaten*“ gilt „*Primärdaten als Grundlagen für Veröffentlichungen sollen auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, zehn Jahre lang aufbewahrt werden*“.¹⁴⁵ Diese Forderung ist nicht beschränkt auf bestimmte Wissenschaftsbereiche. Für die Nachvollziehbarkeit der wissenschaftlichen Ergebnisse ist daher eine Archivierung der Forschungsdaten unerlässlich und stellt gleichzeitig die gute wissenschaftliche Praxis dar. Wie die Archivierung von Primärdaten jeweils durchgeführt wird, obliegt der jeweiligen Institution und ist nicht zentral geregelt. Unter einer digitalen Langzeitarchivierung (LZA) wird jedoch nicht nur eine reine Sicherung der Forschungsdaten auf einem Speichermedium (beispielsweise Festplatten, Bandlaufwerken oder auch langzeitarchivierungsfähige optische Datenträger) verstanden. Vielmehr müssen bei einer digitalen Langzeitarchivierung laut nestor, dem deutschen Kompetenznetzwerk zur digitalen Langzeitarchivierung, alle folgenden Punkte berücksichtigt werden:

- Bitstream-Preservation und Speichermedien
- Erhaltungsplanung
- Formate
- Erhaltungsstrategien
- Metadaten und Struktur

Da es aus Platzgründen nicht möglich ist jeden dieser Punkte hier im Einzelnen zu erläutern, sei auf eine Einführung in die digitale Langzeitarchivierung bei nestor¹⁴⁶ verwiesen. Eine reine Bitstream-Preservation, also der Erhalt der Bits ohne geeignete Erhaltungsstrategien, und eine Archivierung von Forschungsdaten ohne Metadatenmanagement ist für eine sichere Langzeitarchivierung nicht ausreichend, denn ohne Erhaltungsplanung und Metadaten ist die Zugreifbarkeit und sinnvolle Nachnutzbarkeit von Forschungsdaten nicht sichergestellt. Da die technische Umsetzung der Archivierung von komplexen Forschungsdaten bis hin zur Archivierung von kompletten Forschungsumgebungen noch nicht umfänglich gelöst ist, sondern

¹⁴⁴ http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf [letzter Zugriff 1.4.2015].

¹⁴⁵ Sicherung guter wissenschaftlicher Praxis, http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf [letzter Zugriff 1.4.2015].

¹⁴⁶ Nestor Einführung in die LZA, <https://wiki.dnb.de/pages/viewpage.action?pageId=94681401> [letzter Zugriff 13.4.2015].

sich noch immer in der Entwicklung befindet, bewegen sich die später vorgestellten Handlungsvorschläge für einen standardisierten Aufbau von Archiven eher in einem Zeitrahmen von ca. 10–15 anstatt 100 Jahren.

Aktuell wird die Speicherung der Forschungsdaten an den verschiedenen Landesuniversitäten unterschiedlich gehandhabt. Es werden im Folgenden einige universitäre Dienstangebote aufgelistet, wobei solche, die ihren Fokus nur auf bestimmte Fachbereiche oder Dateiformate liegen oder lediglich zur Publikation archivierter Daten nicht näher behandelt werden. Beispiele für solche spezialisierten Angebote stellen die Dienste der GESIS¹⁴⁷ für Sozialwissenschaftliche Umfragedaten oder auch die Archive für die LHC-Experimente innerhalb von GridKA¹⁴⁸ dar. Der Fokus liegt daher auf Diensten, die eine Archivierung oder zumindest eine Speicherung von beliebigen Forschungsdaten ermöglichen. Da, wie später in 5.3.3.2. OAIS-Referenzmodell näher erläutert wird, aktuell eine Archivierung im vollumfassenden Sinn an keinem Standort angeboten wird, gibt diese Auflistung hauptsächlich eine Übersicht von Diensten, die zumindest eine langfristige Speicherung von Forschungsdaten ermöglichen. Falls bekannt, wurden auch sich in Entwicklung befindende Dienste berücksichtigt. Im Anschluss der Diensteübersicht wird ein standardisiertes Referenzmodell vorgestellt, das dazu dienen könnte, geeignete Archive innerhalb von Baden-Württemberg zu entwickeln. Dieses Referenzmodell dient als Grundlage für die darauffolgenden Handlungsvorschläge.

5.3.3.1 Dienste

Freiburg In Freiburg wird seit Mai 2015 die erweiterte Publikationsplattform FreiDok plus¹⁴⁹ angeboten. Im Laufe des Jahres wird der Dienst um eine Hochlademaske erweitert werden, die es den Mitgliedern der Universität Freiburg kostenfrei ermöglicht, Forschungsdaten inklusive Metadaten Open Access zu veröffentlichen. Der Dienst ist nur für die Veröffentlichung von Forschungsdaten vorgesehen, eine reine Ablage ohne Veröffentlichungscharakter ist nicht möglich. Allerdings ist es möglich eine Embargo-Frist einzustellen, um den Zugriff auf die Daten für eine gewisse Dauer zu beschränken. Hierbei haben dann nur die Kuratoren der Bibliothek Zugriff auf die Daten. FreiDok plus wird von der Bibliothek kuratiert und alle Forschungsdaten sind mittels DOI persistent zugreif- sowie zitierbar. Die Daten sollen ohne Befristung, jedoch für mindestens 10 Jahre verfügbar gehalten werden. Eine Löschung von Daten ist aufgrund der persistenten DOI nicht vorgesehen. Eine weitere Speichermöglichkeit in Form eines TSM/ADSM-Archivs wird vom Rechenzentrum der Universität Freiburg angeboten. Kleinere Datenmengen können aktuell kostenlos gespeichert werden, bei großen Datenmengen müssen die Kosten für das Band getragen werden. Aktuell besteht eine unbefristete Aufbewahrungsdauer. Der Zugriff auf die Forschungsdaten im TSM-Archiv kann nur durch eine zuvor benannte Kontaktperson erfolgen. In Partnerschaft mit den Teilnehmern des Backup-Regionalverbundes Freiburg, Heidelberg, Karlsruhe und Mannheim wird eine gegenseitige Spiegelung der jeweiligen Archive vorgenommen.

¹⁴⁷ GESIS – Leibniz-Institut für Sozialwissenschaften, <http://www.gesis.org/> [letzter Zugriff 2.4.2015].

¹⁴⁸ Grid Computing Centre Karlsruhe (GridKA), <http://www.gridka.de/> [letzter Zugriff 2.4.2015].

¹⁴⁹ FreiDok plus, <https://www.freidok.uni-freiburg.de/> [letzter Zugriff 11.05.2015].

Heidelberg In Heidelberg stehen zusätzlich zur Speicherung der Forschungsdaten im TSM/ADSM-Archiv die Repositorien HeiDATA¹⁵⁰ und HeidICON¹⁵¹ den Forschenden zur Verfügung. HeiDATA ermöglicht die Speicherung und Publikation von Forschungsdaten für mindestens 10 Jahre. In HeiDATA können beliebige Dateiformate gespeichert werden. Alle Daten sind auch mittels DOI dauerhaft zugreif- und zitierbar. Im Gegensatz dazu fungiert HeidICON speziell als Bild- und Multimediaarchiv.

Hohenheim Die Universität Hohenheim betreibt in Kooperation mit der Universität Stuttgart ein TSM/ADSM-Archiv und ermöglicht ihren Forschenden daher ihre Forschungsdaten zu speichern. Standardmäßig werden die Daten nach einem Jahr ohne Rückfrage gelöscht. Auf Antrag hin kann die Speicherdauer jedoch verlängert werden.

Karlsruhe An der Universität Karlsruhe können Forschungsdaten ebenfalls mittels TSM/ADSM gespeichert werden. Die Standardspeicherdauer beträgt ein Jahr, kann jedoch auf 90 Tage, 1-4 oder 10 Jahre eingestellt werden. Überdies arbeitet die Universität Karlsruhe in Kooperation mit verschiedenen Partnern an den Projekten bwDataArchiv¹⁵² und RADAR¹⁵³. bwDataArchiv mit der Projektlaufzeit 1.1.2014 – 31.12.2016 fokussiert sich zunächst darauf, die unterste Schicht (Bit-Preservation) anderen Projekten, wie bspw. bwDataDiss¹⁵⁴, zur Verfügung zu stellen.

„Die Entwicklung des neuen Dienstes soll bis 2015 so weit vorgeschritten sein, dass Daten von abgeschlossenen Projekten des HLRS aufgenommen und verlässlich archiviert werden können. Anhand der ersten Erfahrungen soll der Dienst im Projekt für weitere Nutzer etabliert werden und im Hinblick auf Integrität, Nutzbarkeit und Verfügbarkeit im zweiten Schritt weiterentwickelt werden.“ (bwDataArchiv – Wissenschaftliche Datenarchivierung¹⁵⁵)

Daher ist es aktuell nicht vorgesehen Forschungsdaten der einzelnen Forschenden direkt in bwDataArchiv zu archivieren. Das Projekt RADAR entwickelt aktuell ein Research Data Repository, dessen Dienst später deutschlandweit verfügbar sein soll. Zwei Angebote sollen zur Verfügung stehen. Zum einen ein Publikationsangebot, bei der eine dauerhafte Archivierung angestrebt wird. Zum anderen soll aber auch ein Archivierungsangebot bereitgestellt werden, in dem jegliche Forschungsdaten für 5, 10 oder auch 15 Jahre archiviert werden können. Eine genaue Preiskalkulation wird derzeit entwickelt, orientiert sich jedoch u. a. an den unterschiedlichen Haltefristen. Ein halbes Jahr vor Ablauf der Archivierungsfrist wird der Ansprechpartner über weitere Möglichkeiten der Archivierung informiert. Nach Ablauf der

¹⁵⁰ HeiDATA Dataverse Network, <https://heidata.uni-heidelberg.de/dvn/> [letzter Zugriff 2.4.2015].

¹⁵¹ HeidICON – Die Heidelberger Bild- und Multimediadatenbank, <https://heidicon.ub.uni-heidelberg.de/> [letzter Zugriff 2.4.2015].

¹⁵² bwDataArchiv – Wissenschaftliche Datenarchivierung, <http://www.alwr-bw.de/kooperationen/bwdataarchiv/> [letzter Zugriff 7.5.2015].

¹⁵³ RADAR, <http://www.radar-projekt.org/> [letzter Zugriff 7.5.2015].

¹⁵⁴ In bwDataDiss sollen Dissertationen inklusiver der Forschungsdaten archiviert werden.

¹⁵⁵ <http://www.alwr-bw.de/kooperationen/bwdataarchiv/> [letzter Zugriff 7.5.2015].

Frist werden die Daten gelöscht. Abhängig vom gewählten Angebot sind verschiedene Zugriffsrechte geplant. Die Metadaten können (1) öffentlich, (2) für freigegebene Nutzer sowie (3) nicht suchbar / sichtbar sein. Beim reinen Archivierungsangebot sind Forschungsdaten standardmäßig nicht sichtbar, können jedoch für einen beschränkten Kreis an Nutzern durch den Datenkurator freigegeben werden. Publierte Forschungsdaten sollen mittels persistenten Identifiern langfristig zugreif- und zitierbar sein.

Mannheim Die Universität Mannheim bietet über ihren TSM/ADSM-Dienst keine Speicherung von Forschungsdaten an. Allerdings steht den Angehörigen der Universität das Repositorium MADATA¹⁵⁶ zur Verfügung. Wie bei den anderen institutionellen Repositorien können beliebige Datenformate gespeichert werden und sind mittels DOI langfristig zugreif- und zitierbar. Der Zugriff kann über verschiedene Creative-Commons-Lizenzen bis hin zu einem Zugriff nur für ausgewählte Personen geregelt werden.

Stuttgart Ebenfalls über TSM/ADSM können die Forschungsdaten an der Universität Stuttgart für eine bisher unbefristete Aufbewahrungsdauer gespeichert werden. Das angebotene HPSS-Datenmanagement ist für die Archivierung von Forschungsdaten nicht geeignet, da hier lediglich auf Antrag, Speicherplatz für Computing bereitgestellt wird.

Tübingen In Tübingen steht ein Archivierungsdienst zur Verfügung, der es jedem Forschenden ermöglicht kostenlos 1 GB zu speichern. Darüber hinausgehende Datenmengen müssen mit 0,02 € pro GB im Monat bezahlt werden. Die Daten werden ohne zeitliche Befristung und mittels Georeplikation sicher gespeichert. Gelöscht werden die Daten lediglich beim Ausscheiden des Forschers bzw. wenn die Daten durch den Forschenden selbstständig gelöscht werden. Ein öffentlicher Zugriff auf die Daten ist nicht vorgesehen. Dieser Dienst ist nicht für die Archivierung im eigentlichen Sinne geeignet, da das Metadatenmanagement durch den Nutzer erfolgt. Um diesen Umstand zu begegnen, wird aktuell ein Repositorium mit Schwerpunkt Geisteswissenschaften (Digital Humanities) aufgebaut, das laut eigenen Angaben den derzeitigen Vorstellungen für ein Forschungsdatenmanagement entspricht. Das Projekt befindet sich aktuell in der Prototypenphase, weshalb ein Zeitpunkt für einen allgemein verfügbaren Zugang noch nicht feststeht. Das Repositorium basiert auf dem Open-Source-Tool Dokupack und soll beim Archivieren die Forschungsdaten mit verbindlichen Metadaten versehen sowie durch ein weiteres Tool eine Konvertierung der Daten durchführen, um dem Problem der nicht standardisierten Dateiformate zu begegnen und eine Nachnutzung zu erleichtern. Die Originaldatei soll jedoch stets parallel vorhanden bleiben und gemeinsam nach einer Validierung im Archiv abgelegt werden. Die Forschungsdaten sollen mittels persistenten Identifikatoren zitierbar sein, wobei der Forscher definieren können soll, ob die Daten öffentlich zur Verfügung stehen bzw. wer diese Daten sehen und nutzen darf.¹⁵⁷

¹⁵⁶ MADATA – Mannheim Research Data Repository, <https://madata.bib.uni-mannheim.de/> [letzter Zugriff 2.4.2015].

¹⁵⁷ Aufbau eines Archivsystems für die Geisteswissenschaften, <http://www.escience.uni-tuebingen.de/infrastruktur.html> [letzter Zugriff 07.05.2015].

Konstanz / Ulm Die Standorte Konstanz und Ulm bieten aktuell keine zentralen Dienste an, die für eine langfristige Archivierung oder Speicherung von Forschungsdaten geeignet sind.

Aktuell wird an den Universitäten hauptsächlich eine Form der Bitstream-Preservation durchgeführt. Gerade bei den TSM-Archiven ist meistens nicht klar, welche Daten auf den Bändern gelagert sind, da kein zentrales Metadatenmanagement stattfindet. Dies birgt die Gefahr, dass sogenannte Datengräber entstehen, da meistens nicht klar ist, welche Daten sich in den Archiven befinden und wie diese nachgenutzt werden können. Dass das TSM nicht allein für die Langzeitarchivierung geeignet ist, zeigt auch das Zitat aus der Stuttgarter Dienstbeschreibung:

„Beachten Sie bitte, dass die Archive-Funktion mit TSM, wie wir sie anbieten können, wegen ihrer Plattformgebundenheit keine Langzeitarchivierung, die gesetzlichen Vorgaben entsprechen würde, darstellt!“ (FAQs zu TSM¹⁵⁸)

Aber auch Repositorien sind ohne Erhaltungsplanung, Risikomanagement, usw. nicht als Langzeitarchive vollumfänglich geeignet, da die Forschungsdaten zwar möglicherweise mittels Metadaten gefunden werden können und zugreifbar sind, eine Nachnutzung allerdings nicht sichergestellt werden kann, falls beispielsweise veraltete, möglicherweise proprietäre Formate verwendet wurden. Repositorien können jedoch dazu beitragen, die Komplexität der Archivierung und Veröffentlichung an einer zentralen Stelle zusammenzufassen. Forschende könnten daher davon profitieren, wenn die Veröffentlichung von Forschungsdaten und eine darin inbegriffene Archivierung Hand in Hand stattfindet. Jedoch ist nicht immer eine Veröffentlichung von allen Forschungsdaten für die Allgemeinheit vorgesehen. Dieser Aspekt sollte bei der Planung von Archiven berücksichtigt werden. Für den Fall, dass Repositorien zur Archivierung von Forschungsdaten genutzt werden sollen, muss entweder ein eigenständiges Archiv entwickelt werden, das von den Repositorien verwendet werden kann, oder die Repositorien müssen zu geeigneten Archiven weiterentwickelt werden.

5.3.3.2 OAIS-Referenzmodell

Um die Praxis der digitalen Langzeitarchivierung zu standardisieren, wurde das Open Archival Information System (OAIS) entwickelt. Das OAIS-Modell behandelt sowohl technische als auch Metadaten- und Rollen Aspekte und zielt auf die dauerhafte Aufbewahrung und die Gewährleistung einer langfristigen Zugänglichkeit digitaler Objekte ab. Im Jahr 2003 wurde das OAIS-Modell in den ISO Standard 14721 überführt und als aktualisierte Version 2 im Jahr 2012 in ISO 14721:2012 übernommen. Da das OAIS-Modell allgemeingültig ist und somit keine Beschränkungen auf bestimmte Datenformate oder Systemarchitekturen besitzt, stellt dieses Modell, gerade in einem heterogenen Universitätsumfeld, eine gute Grundlage der Archivierung dar.

¹⁵⁸ http://www.tik.uni-stuttgart.de/dienste/datensicherung/backup/TSM_FAQs.pdf [letzter Zugriff 11.5.2015].

Da aus Platzgründen lediglich eine kurze Beschreibung des Referenzmodells möglich ist, sei für eine detaillierte Beschreibung auf die deutsche Übersetzung durch nestor¹⁵⁹ verwiesen. Das OAIS-Referenzmodell (Abb. 39) unterscheidet zwischen den drei Informationspaketen SIP, AIP und DIP. Digitale Daten, die in das Archiv übernommen werden sollen (Ingest), werden Submission Information Packages (SIP) genannt. Diese SIP werden durch archivarische Metadaten (Descriptive Info) angereichert und gelten dann als Archival Information Packages (AIP), die archiviert (Archival Storage) werden. Beim Zugriff (Access) erfolgt eine Generierung der Daten in Dissemination Information Packages (DIP), wobei die Daten auf die Rechte und Bedürfnisse der Konsumenten zugeschnitten sind. Somit kann sichergestellt werden, dass Daten mit eingeschränkten Rechten nur an berechnigte Personen sowie bei Bedarf in einem gewünschten Format ausgeliefert werden.

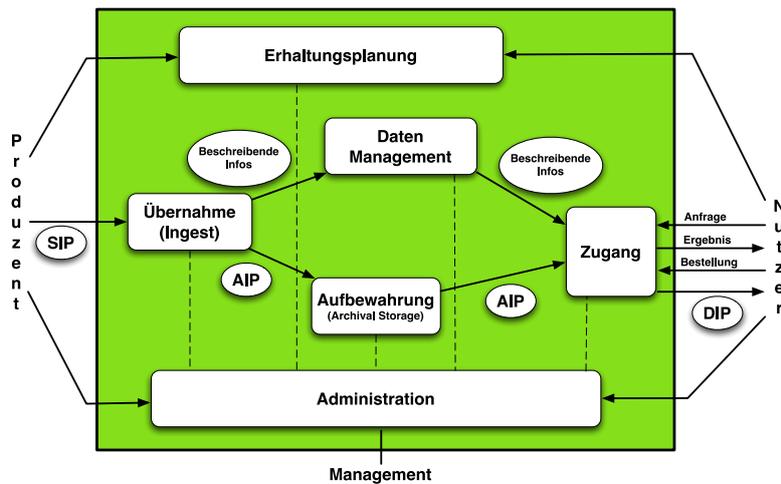


Abb. 39: OAIS-Referenzmodell

Für die Archivierungsvorbereitung ist eine Qualitätssicherung der Daten von zentraler Bedeutung. Dies bedeutet, dass Daten vor dem Einlagern auf ihre Konformität mit einem Standard geprüft werden sollten. Im OAIS-Modell wird die langfristige Zugreifbarkeit durch eine Erhaltungsplanung abgesichert, bei der geplant werden muss, zu welchen Zeitpunkten oder Ereignissen welche Maßnahmen getroffen werden müssen, um die Zugreifbarkeit zu gewährleisten. Außerdem sorgen klar verteilte Rollen und ein kontrollierendes Management für die Einhaltung notwendiger Rahmenbedingungen.

5.3.4 E-Science-Bestrebungen

Innerhalb des Kapitels „AG 4 – Forschungsdatenmanagement“ des ESFK heißt es:

„Während temporäre Zwischenergebnisse verworfen werden, müssen Primärdaten und finale Ergebnisse, die Grundlage wissenschaftlicher Publikationen sind, im

¹⁵⁹ Referenzmodell für ein Offenes Archiv-Informationssystem – Deutsche Übersetzung (nestor-Materialien 16, Version 2), <http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0008-2013082706> [letzter Zugriff 7.5.2015].

Sinne „guter wissenschaftlicher Praxis“ und nach den Richtlinien der Fördermittelgeber nach definierten Standards mit Metadaten dokumentiert und nachhaltig gesichert werden.“ (ESFK, 82)

Hierdurch wird die Wichtigkeit der langfristigen Archivierung dokumentiert. Wie wichtig die Dokumentation und Metadatenanreicherung im Hinblick auf die Archivierung und Nachnutzung ist, zeigt die weitere Aussage:

„Die Daten müssen so dokumentiert sein, dass der Nutzer verstehen und nachvollziehen kann, wie und wofür sie erzeugt wurden. Erst dann kann man davon ausgehen, dass auch andere Nutzer Interesse haben, die Daten für ihre Projekte zu verwenden.“ (ESFK, 84)

Es wird auch angemerkt, dass eine reine Datenspeicherung nicht ausreichend ist, sondern eine langfristige Zugriffssicherung notwendig sein kann:

„Im Hinblick auf die Nachnutzer müssen Datenformate gewählt werden, deren Strukturen vollständig offengelegt sind, damit sie von zukünftigen Forschenden interpretiert werden können. Proprietäre Formate stellen ein Risiko für die spätere Interpretierbarkeit der Daten dar. Falls sie unumgänglich sind, müssen Konvertierungsroutinen und Umgebungen (Emulatoren), wie sie z. B. im Landesprojekt „Funktionale Langzeitarchivierung“ (bwFLA) untersucht werden, vorgehalten und gesichert werden.“ (ESFK, 84)

Aus diesen Gründen werden als Maßnahmen für die Hochschulen in Baden-Württemberg folgende Aspekte des Förderprogramms dargelegt¹⁶⁰: (1) die Entwicklung eines Archivierungsworkflows für typische Datenarten, (2) der Umgang mit proprietären, nicht dokumentierten Dateiformaten, (3) die Entwicklung eines Geschäftsmodells, das sich aus Kosten für Storage und formatspezifischen Zusatzpauschalen (z. B. für Emulatoren) zusammensetzen, und (4) wirtschaftliche und rechtliche Überlegungen (z. B. Haftung, Eigentümerschaft, Persönlichkeitsschutz) anstellen. Unter der Einrichtung einer landesweiten Koordination wird überdies eine „Landesweite Abstimmung und Umsetzung von Policies, Methoden und Verrechnungsmodelle für Langzeitarchivierung“¹⁶¹ skizziert. Die genannten Punkte sind für die digitale Langzeitarchivierung sinnvoll und notwendig, jedoch gibt es bisher wenig Umsetzung dieser geforderten Punkte. Teilweise werden diese Punkte, wie in 5.3.3. [Status quo](#) aufgezeigt, aber bereits angegangen.

5.3.5 Gegenwärtige Handlungsempfehlungen

Da eine Langzeitarchivierung von Forschungsdaten innerhalb der Landesuniversitäten bisher kaum gegeben ist, bedarf es der Schaffung eines funktionierenden Systems für die Archivierung. Hierfür werden jedoch nicht nur Speichersysteme sondern auch Workflows und Richtlinien benötigt.

¹⁶⁰ ESFK, 89.

¹⁶¹ ESFK, 90.

Systeme für die Archivierung Für den Aufbau eines Archivierungssystems eignet sich das OAIS-Referenzmodell. Hierzu gehört jedoch nicht nur der Aufbau einer Infrastruktur, sondern eine klare Zuweisung aller Rollen des Referenzmodells. Insbesondere werden klare Verantwortlichkeiten für das Datenmanagement und etablierte Standardworkflows benötigt. Die Etablierung von FDM-Experten könnte erheblich helfen, die Datenaufbereitung in Absprache mit den Forschenden zu steuern.

Workflows für die Archivierung Weitere wichtige Bausteine sind noch zu definierende Workflows für die Archivierung. Wünschenswert und notwendig sind vor allem allgemein verständliche Vorgaben, an die sich die Forschenden orientieren können und die beschreiben, welche Schritte für eine nachhaltige Archivierung notwendig sind. Beispielhaft könnte ein Modell folgende Punkte umfassen: (1) Informationsangebote, (2) Hilfs- und Beratungsstellen sowie (3) Abgabe der Forschungsdaten an einer definierten Stelle. (1) beinhaltet die Forschenden zu informieren, welche Datenformate für eine Langzeitarchivierung geeignet sind, welche Daten Archivierungswürdig sind und wie ein möglicher Workflow (samt Rechteübergabe) aussieht. Bei weiteren Fragen oder Problemen müssen Hilfs- und Beratungsstellen (2) zur Verfügung stehen. Zuletzt muss nach dem Referenzmodell eine definierte Abgabestelle (3) eine Qualitätssicherung durchführen. Das könnte beispielsweise durch Forschungsdatenbeauftragte (s. [3.1.4. Gegenwärtige Handlungsempfehlungen](#)) geschehen.

Datenmanagementpläne (DMPs) Abgesehen von einem Archivierungssystem und notwendigen Workflows zur Archivierung, stellt die Langzeitarchivierung jedoch nur den letzten Punkt des Datenmanagements dar. Ohne Datenmanagementpläne ist eine Archivierung kaum möglich. Datenmanagementpläne sind wichtig, um mögliche Nachnutzungsszenarien zu definieren. So können Daten schon bei der Erzeugung geeignet dokumentiert werden. Ohne DMPs ist Archivierung oft ineffizient und unverhältnismäßig kostspielig.

Langfristige Zugriffssicherung Für eine langfristige Zugriffssicherung bedarf es entsprechender Dienste, um mittels eines einfachen Workflows Daten samt Software und Systemen zu archivieren.

Finanzierung Das Kernproblem der Finanzierung sollte durch DMPs bereits gelöst werden, da zum einen im Voraus geplant werden muss, wie und wo die Daten aufbewahrt werden sollen und die dafür notwendige Finanzierung beantragt werden sollte. Dem zweiten Problem der projektunabhängigen Speicherung von Forschungsdaten kann begegnet werden, indem ein kostenloser Speicher für die Langzeitarchivierung (mindestens nach DFG-Richtlinien) kleiner Datenmengen überall verfügbar wird.

Allgemein Abseits der genannten Probleme muss sich verstärkt mit der Rechtsklärung bezüglich der Softwarearchivierung befasst werden. Um archivierte Forschungsdaten funktional und somit langfristig zugreifbar nachnutzen zu können, müssen sowohl die notwendigen Betriebssysteme als auch die Fachsoftware archiviert werden. Um dies zu ermöglichen müssen jedoch lizenzrechtliche Probleme geklärt werden.

Zusammenfassend lässt sich festhalten, dass die aktuelle Praxis der reinen Bitstream-Erhaltung an den Universitäten für eine zukunftsichere Nachnutzung nicht ausreichend ist. Um diese zu gewährleisten, sollte an den Universitäten ein vollumfassendes Gesamtkonzept nach OAIS-Standard, bestehend aus Beratung und Qualitätssicherung, entwickelt werden. Hierfür sind die darin genannten Rollen, Workflows und Verantwortlichkeiten festzulegen sowie klare und transparente Definitionen und Hilfestellungen bezüglich einzuhaltender Standards herauszuarbeiten.

6 Weitere Aspekte

6.1 IT-Infrastruktur (445)

6.1.1 Einleitung

In der modernen Wissenschaft werden ein Großteil der Daten digital aufbewahrt und bearbeitet. Effizientes Forschungsdatenmanagement setzt das Vorhandensein von entsprechenden Speicher- und Rechenkapazitäten voraus. Wachsende Datenmengen erfordern außerdem schnelle Datenleitungen. Sicherheit von Forschungsdaten hängt im hohen Maße von der Zuverlässigkeit der angewendeten Speichersysteme sowie von Qualität und Umfang von Sicherungslösungen ab. Damit bildet die Effizienz und die Leistungsfähigkeit der zur Verfügung stehenden IT-Infrastruktur eine Voraussetzung für ein effizientes Forschungsdatenmanagement.

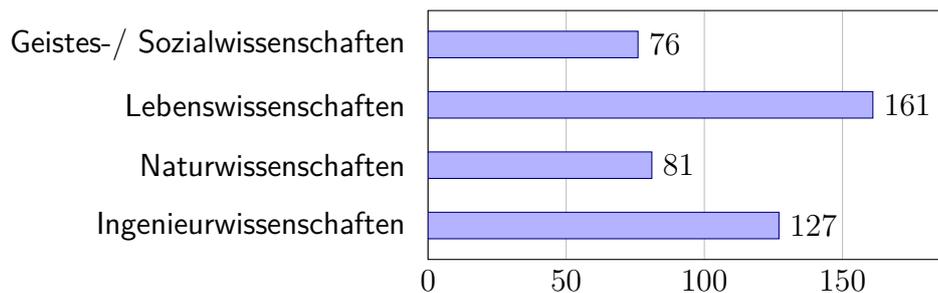


Abb. 40: Verteilung der User-Stories nach der Fachsystematik der DFG

Das Kapitel IT-Infrastruktur beschreibt die Bereiche: Geräteausstattung, Rechenleistung / Cluster / HPC, virtuelle Maschinen sowie Datenspeicherung, -zugriff und -sicherheit. Die gesamte Anzahl von User-Stories, auf welchen dieses Kapitel basiert, beträgt 445 (Abb. 40). Aus dem Bereich Lebenswissenschaften stammt davon mehr als ein Drittel (161). Aber auch die Ingenieurwissenschaften sind mit 127 User-Stories verhältnismäßig stark vertreten. Von den Forschenden der Geistes- und Sozialwissenschaften sowie der Naturwissenschaften wurden jeweils etwa 80 Wünsche geäußert. Die Anforderungen der Geisteswissenschaften, vor allem relativ zu ihrer Stichprobenhäufigkeit, sind im Vergleich zu den anderen Wissenschaftsbereichen signifikant geringer.

6.1.2 Bedarf der Wissenschaftsgemeinschaft

6.1.2.1 Geräteausstattung (56)

„Als Forschende, welche für die Durchführung eigener Studien auf moderne Geräteausstattung angewiesen sind, wünschen wir uns eine Erweiterung unserer Ausstattung und /oder einen vereinfachten Antragsweg für die Beschaffung von Geräten, um ohne Einschränkungen durch veraltete oder nicht vorhandene Geräte forschen zu können.“

Ein Teil der Forschenden gab an, dass die Ausrüstung ihrer Forschungsrichtungen mit konkreten Geräten (14) nötig sei. Hinzu kommen allgemeine Wünsche nach besserer Ausstattung (6) und der Wunsch nach einer Verbesserung der Finanzierung, um Ausstattung aufzurüsten (17).

Darüber hinaus wurde der Wunsch geäußert, den Verwaltungsaufwand für die Beschaffung von Geräten zu verringern. Dabei wurde Bedarf nach mehr Selbstständigkeit bei der Beschaffung von notwendigen Ausstattungskomponenten angesprochen (1), nach mehr Flexibilität bei der Beschaffung (4), nach Möglichkeit Beratung bei der Beschaffung zu erhalten (4) sowie nach Beschleunigung und Vereinfachung der Beschaffungsabläufe (3).

6.1.2.2 Mehr Rechenleistung / Cluster / HPC (117)

„Als Forschende, die komplexe Datenanalysen und Simulationen durchführen und genügend Rechenleistung in der Forschungsarbeit benötigen, wünschen wir uns Zugang zu mehr Rechenleistung, um unsere Auswertungen schnell und effizient ausführen zu können nicht durch langsame Hardware ausgebremst zu werden.“

Um zeitlich effizient ihre Arbeit ausführen zu können, wurde von vielen Forschenden der Bedarf nach mehr Rechenleistung in verschiedenen Varianten geäußert. Oft wurde eine leistungsfähigere bzw. modernere Hardware mit leistungsfähigeren Prozessoren, mehr Arbeitsspeicher, größeren Festplatten abseits des HPC-Bereich gewünscht (38).

In den meisten Fällen betraf der geäußerte Bedarf hingegen Großrechner, also den Bereich des High Performance Computing (HPC) (72): Ein schnellerer, einfacherer, flexiblerer Zugang zu HPC-Systemen (26), eine höhere Benutzerfreundlichkeit (5), Stabilität (4), mehr Rechenleistung der HPC-Systeme (16) und eine schnellere Anbindung zu den HPC-Systemen (9), die Verfügbarkeit von bestimmten Softwareprodukten auf den HPC-Rechnern (5) und Workshops und Schulungen zum Thema HPC (3).

Darüber hinaus wurde auch mehrmals der Bedarf nach lokalen Rechenclustern ausgesprochen, um einfacher und ohne Wartezeit eigene Berechnungen durchführen zu können (11).

6.1.2.3 Datenspeicherung, -zugriff und -sicherheit (307)

„Als Forschende, wünschen wir uns Speichersysteme mit speziellen Eigenschaften für unsere Daten, um die stetig wachsenden Menge der Forschungsdaten sinnvoll, sicher und langfristig speichern zu können.“

Das Wachstum der Speicherkapazitäten von Festplatten verläuft weitestgehend entsprechend dem Mooreschen-Gesetz¹⁶². Auch der Speicherbedarf der Forschenden steigt nach ihren Angaben jedes Jahr deutlich (s. Abb. 41).

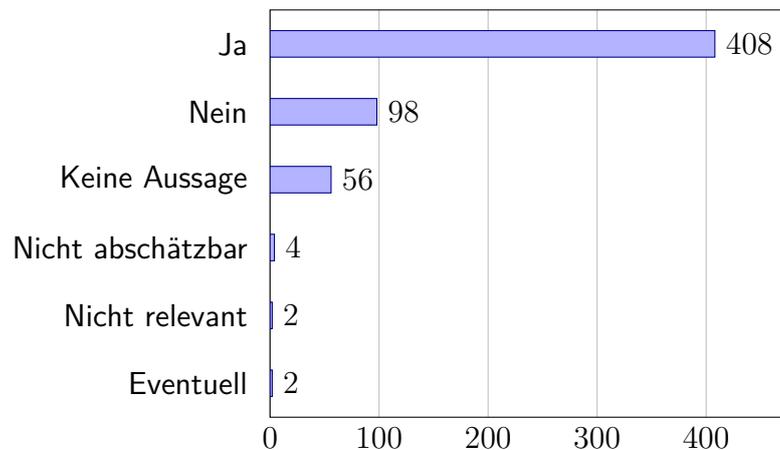


Abb. 41: „Könnte Ihre Datenmenge in Zukunft wachsen?“ (Frage 5)

Die meisten Forschenden thematisieren hierbei Mängel an ausreichender Speicherkapazität. In einigen Fällen stand aber nicht die Größe des Speichers im Mittelpunkt. In solchen User-Stories wurde angesprochen, dass sich die Forschenden für ihre Datenspeicherung eine schnelle Anbindung (51), Zuverlässigkeit und Stabilität (10) sowie Benutzerfreundlichkeit (11) wünschen.

Weiterhin wurde die Bedeutung der Zugriffssteuerung hervorgehoben (52) sowie das Thema Backup thematisiert (87).

6.1.2.4 Virtuelle Maschinen (7)

„Als Forschende, wünschen wir uns den Einsatz von virtuellen Maschinen, um flexibler mit Software arbeiten zu können.“

Eine virtuelle Maschine¹⁶³ ist die Nachbildung eines real existierenden oder hypothetischen Rechnersystems. Auf einer virtuellen Maschine kann eine Software oder ein ganzes Betriebssystem wie auf einem physischen Rechner installiert und genutzt werden. Das ermöglicht beispielsweise den parallelen Einsatz verschiedener Betriebssysteme auf einem Rechner und dadurch eine flexible Nutzung von vorhandenen Ressourcen. Zum Einsatz von virtuellen Maschinen wurden nur relativ wenige Wünsche geäußert. Das mag auch daran liegen, dass virtuelle Maschinen eher ein Mittel als ein Zweck sind und trotz vieler Anwendungen nur von IT-Experten explizit gewünscht werden.

¹⁶² U. Cayoglu, D. Shcherbakov, F. Tristram, Research Data Management within the e-Science Strategy in Baden-Württemberg. Preliminary Results of Project bwFDM-Communities, 2nd Data Management Workshop, Cologne, November 28th -29th, 2014.

¹⁶³ G.J. Popek, R.P. Goldberg, Formal Requirements for Virtualizable Third Generation Architectures. In: Communications of the ACM. 17, Nr. 7, 1974, 412-421. doi:10.1145/361011.361073 [letzter Zugriff 29. 6. 2015].

6.1.3 Status quo

6.1.3.1 Rechenleistung/Cluster/HPC

Standort	Bezeichnung	Fachrichtung
HLRS - Höchstleistungsrechenzentrum Stuttgart ¹	Hornet ²	antragsabhängig
Freiburg (geplant Q3.2015)	bwForCluster ENM	Elementarteilchenphysik, Neurowissenschaften, Mikro- systemtechnik
Tübingen (geplant Q3.2015)	bwForCluster BinAC	Bioinformatik und Astrophy- sik
Ulm(aktiv seit 12/2014)	bwForCluster - Justus ³	Theoretische Chemie
Karlsruhe	bwUniCluster ⁴	Allgemeine Fachbereichsver- sorgung, Grundversorgung und Lehre
Karlsruhe	ForHLR ⁵	antragsabhängig
Esslingen	bwGRiDCluster ⁶	Probleme aus Wissenschaft und Wirtschaft

¹ Hornet: <http://www.hlrs.de/> [letzter Zugriff 17.6.2015]

² Hornet - aktueller Rang 16 in Top500: <http://www.top500.org/lists/2014/11/> [letzter Zugriff 17.6.2015]

³ bwForCluster-Justus: <http://www.uni-ulm.de/einrichtungen/kiz/it/rechner-compute-server/justus.html> [letzter Zugriff 17.6.2015]

⁴ bwUniCluster: <http://www.scc.kit.edu/dienste/bwUniCluster.php> [letzter Zugriff 17.6.2015]

⁵ ForHLR: <https://www.scc.kit.edu/dienste/forhrlr.php> [letzter Zugriff 17.6.2015]

⁶ bwGRiDCluster <http://www.bw-grid.de/das-bwgrid/verbliebene-standorte> [letzter Zugriff 17.6.2015]

Tab. 7: Zugangsangebote zu den HPC-Clustern in Baden-Württemberg.

Aktuelle Zugangsangebote zu den HPC-Clustern in Baden-Württemberg sind in [Tab. 7](#) dargestellt. Die Entwicklung des Hochleistungsrechnens in Baden-Württemberg läuft im Rahmen des bwHPC Konzeptes¹⁶⁴. Es werden HPC-Lösungen für verschiedene Forschungsrichtungen mit unterschiedlichen Leistungsanforderungen zur Verfügung gestellt. Die höchste Leistungsstufe in Baden-Württemberg repräsentiert das HPC-System Hornet (Cray XC40) im Höchstleistungsrechenzentrum Stuttgart¹⁶⁵. In der aktuellen Liste der schnellsten Rechner der Welt steht dieses HPC-System mit der Spitzeleistung von 3784,3 TFlop/s auf Platz 16¹⁶⁶. Der

¹⁶⁴ H. Hartenstein, T. Walter, P. Castellaz, Aktuelle Umsetzungskonzepte der Universitäten des Landes Baden-Württemberg für Hochleistungsrechner und datenintensive Dienste. PIK - Praxis der Informationsverarbeitung und Kommunikation, 36(2), 2013, 99-108. doi:10.1515/pik-2013-0007 [letzter Zugriff 16.5.2015]; Baden-Württembergs Umsetzungskonzept für das Hochleistungsrechner http://www.bwhpc-c5.de/bwhpc_konzept.php [letzter Zugriff 30.6.2015].

¹⁶⁵ HLRS - Höchstleistungsrechenzentrum Stuttgart <http://www.hlrs.de/> [letzter Zugriff 30.6.2015].

¹⁶⁶ TOP500 Supercomputer <http://www.top500.org/lists/2014/11/> [letzter Zugriff 17.6.2015]

HPC-Cluster Justus¹⁶⁷ wurde in der TOP500 ebenso aufgelistet, mit der Spitzenleistung von 272,8 TFlop/s, Platz 275¹⁶⁸ der Weltrangliste.

Auf der Webseite des Begleitprojektes bwHPC-C5¹⁶⁹ befinden sich umfangreiche Informationen zu der Nutzung¹⁷⁰ und dem Zugang zu den HPC-Ressourcen in Baden-Württemberg. Für grössere Rechenvorhaben wird spezielles Knowhow in den sogenannten Tiger-Teams zur Verfügung gestellt¹⁷¹. Es wurden auch Schulungsangebote aufgebaut¹⁷². Ebenso wurde im Rahmen des bwHPC-C5-Projektes ein bwHPC-Wiki eingerichtet¹⁷³, welches kontinuierlich erweitert und aktualisiert wird.

6.1.3.2 Datenspeicherung, -zugriff und -sicherheit

Die Tabellen (Tab. 8, Tab. 9, Tab. 10 und Tab. 11) geben einen Überblick zu den Datenspeicherangeboten an den Landesuniversitäten.

¹⁶⁷ bwForCluster-Justus <http://www.uni-ulm.de/einrichtungen/kiz/it/rechner-computer-server/justus.html> [letzter Zugriff 30.6.2015].

¹⁶⁸ TOP500 Supercomputer <http://www.top500.org/lists/2014/11/> [letzter Zugriff 17.6.2015]

¹⁶⁹ bwHPC-C5 - Projekt zur Koordination einer hochschulübergreifenden Unterstützung für die Nutzer der baden-württembergischen Hochleistungsrechner <http://www.bwhpc-c5.de/> [letzter Zugriff 30.6.2015].

¹⁷⁰ Informationen zur Nutzung von HPC-Systemen in Baden-Württemberg <http://www.bwhpc-c5.de/nutzerunterstuetzung.php> [letzter Zugriff 30.6.2015].

¹⁷¹ Gezielte Anwenderunterstützung durch Tiger-Teams <http://www.bwhpc-c5.de/nutzerunterstuetzung.php> [letzter Zugriff 30.6.2015].

¹⁷² Schlingen zum Thema HPC <http://java.hlrs.de/bwHPC-Veranstaltungen/> [letzter Zugriff 30.6.2015].

¹⁷³ bwHPC-Wiki http://www.bwhpc-c5.de/wiki/index.php/Main_Page [letzter Zugriff 30.6.2015].

Dienst	Standort	Nutzer
Cloudspeicher bwSync&Share ^{1,2} bwFileStorage ³	KIT, Karlsruhe Mannheim/Heidelberg	Mitarbeiter und Studierende der Landesuniversitäten HPC-Nutzer bekommen Zugang zu Large Scale Data Facility (LSDF) ⁴
Fileserver	Jede Universität	Mitarbeiter und Studierende der Landesuniversitäten
Backup (TSM) ⁵ ILIAS ⁶	Jede Universität Jede Universität	Mitarbeiter und Studierende der Landesuniversitäten Mitarbeiter und Studierende der Landesuniversitäten

¹ Info zu 'bwSync&Share': <http://www.alwr-bw.de/kooperationen/bwsync-share/> [letzter Zugriff 30.6.2015]

² Homepage 'bwSync&Share': <https://bwsyncandshare.lsd.f.kit.edu/login> [letzter Zugriff 30.6.2015]

³ 'bwFileStorage': <http://www.alwr-bw.de/kooperationen/bwfilestorage/> [letzter Zugriff 30.6.2015]

⁴ 'Large Scale Data Facility (LSDF)': <http://www.alwr-bw.de/kooperationen/bwlsdf/> [letzter Zugriff 30.6.2015]

⁵ TSM Tivoli Storage Manager: <http://www-03.ibm.com/software/products/de/tivostormana> [letzter Zugriff 30.6.2015]

⁶ ILIAS - Integriertes Lern-, Informations- und Arbeitskooperationssystem: www.ilias.de [letzter Zugriff 30.6.2015]

Tab. 8: Datenspeicherangebote an den Universitäten in Baden-Württemberg.

Standort	Speicherkapazität	Erweiterungsmöglichkeit
Freiburg	2 Gb ¹	Erweiterbar in Ausnahmefällen ¹
Heidelberg	2 Gb ²	Geringfügig erweiterbar
Hohenheim	2 Gb ³	Geringfügig erweiterbar
Karlsruhe	10 Gb ⁴	-
Konstanz	15 Gb ⁵	-
Mannheim	200 Mb ⁶	Erweiterung möglich
Stuttgart	nicht angegeben ⁷	-
Tübingen	5 Gb ⁸	Nicht erweiterbar
Ulm	100 Mb (Unix und Linuxsysteme), 300 Mb (PC-Pools Windows) ⁹	Für Linux und Unixsysteme erweiterbar auf 150 Mb

¹ Info zu Freiburg:

<http://www.rz.uni-freiburg.de/services/serverdienste/fileserver> [letzter Zugriff 17.6.2015]

² Info zu Heidelberg:

<http://www.urz.uni-heidelberg.de/datenhaltung/user.html> [letzter Zugriff 17.6.2015]

³ Info zu Hohenheim:

<https://kim.uni-hohenheim.de/95428> [letzter Zugriff 17.6.2015]

⁴ Info zu Karlsruhe (KIT):

<https://www.scc.kit.edu/dienste/8639.php> [letzter Zugriff 17.6.2015]

⁵ Info zu Konstanz:

<https://www.rz.uni-konstanz.de/dienste/files/> [letzter Zugriff 17.6.2015]

⁶ Info zu Mannheim:

http://www.uni-mannheim.de/rum/ueber_uns/arbeitsgruppen/zs/fileservice_intro/ [letzter Zugriff 17.6.2015]

⁷ Info zu Stuttgart:

http://www.tik.uni-stuttgart.de/dienste/Fileservice_fuer_Institute/ [letzter Zugriff 17.6.2015]

⁸ Info zu Tübingen:

<http://www.uni-tuebingen.de/en/facilities/zentrum-fuer-datenverarbeitung/dienstleistungen/storage/zugang-und-kosten.html#c4740> [letzter Zugriff 17.6.2015]

⁹ Info zu Ulm:

<http://www.uni-ulm.de/einrichtungen/kiz/it/datenhaltung/datenkontingente.html> [letzter Zugriff 17.6.2015]

Tab. 9: Umfang des individuellen Speicherplatzes pro Mitarbeiter.

Standort	Speicherkapazität	Erweiterungsmöglichkeit
Freiburg	nicht angegeben ¹	nicht angegeben ¹
Heidelberg	2 Gb ²	Geringfügig erweiterbar-
Hohenheim	2 Gb ³	Mehr auf Antrag
Karlsruhe	nicht angegeben ⁴	Erweiterung möglich auf Antrag in 100 Gb-Schritten
Konstanz	50 Gb ⁵	Erweiterung möglich
Mannheim	200 Mb ⁶	Erweiterung möglich
Stuttgart	nicht angegeben ⁷	-
Tübingen	50 Gb ⁸	Erweiterung möglich
Ulm	nicht angegeben ⁹	nicht angegeben

¹ Info zu Freiburg:

<http://www.rz.uni-freiburg.de/services/serverdienste/fileserver> [letzter Zugriff 17.6.2015]

² Info zu Heidelberg:

<http://www.urz.uni-heidelberg.de/datenhaltung/user.html> [letzter Zugriff 17.6.2015]

³ Info zu Hohenheim:

<https://kim.uni-hohenheim.de/95493> [letzter Zugriff 17.6.2015]

⁴ Info zu Karlsruhe (KIT):

<https://www.scc.kit.edu/dienste/8737.php> [letzter Zugriff 17.6.2015]

⁵ Info zu Konstanz:

<https://www.rz.uni-konstanz.de/dienste/files/> [letzter Zugriff 17.6.2015]

⁶ Info zu Mannheim:

http://www.uni-mannheim.de/rum/ueber_uns/arbeitsgruppen/zs/fileservice_intro/ [letzter Zugriff 17.6.2015]

⁷ Info zu Stuttgart:

http://www.tik.uni-stuttgart.de/dienste/Fileservice_fuer_Institute/ [letzter Zugriff 17.6.2015]

⁸ Info zu Tübingen:

<http://www.uni-tuebingen.de/en/facilities/zentrum-fuer-datenverarbeitung/dienstleistungen/storage/zugang-und-kosten.html#c4740> [letzter Zugriff 17.6.2015]

⁹ Info zu Ulm:

<http://www.uni-ulm.de/einrichtungen/kiz/it/datenhaltung.html> [letzter Zugriff 17.6.2015]

Tab. 10: Umfang des Speicherplatzes pro Arbeitsgruppe.

Standort	Dienst	Umfang
Freiburg	Backup (TSM) ^{1 2}	Arbeitsplatzrechner
Heidelberg	Backup (TSM) ³	vollständige Systeme
Hohenheim	Backup (TSM) ⁴	Fileserver, Desktoprechner und Notebooks
Karlsruhe	Backup (TSM) ⁵	Fileserver, Desktoprechner und Notebooks
Konstanz	Backup (TSM) ⁶	Dateien von einzelnen Nutzern sowie Arbeitsgruppen
Mannheim	Backup (TSM) ⁷	Server- und Arbeitsrechner
Stuttgart	Backup (TSM) ⁸	Server- und Arbeitsrechner
Tübingen	Backup (TSM) ⁹	Server- und Arbeitsrechner
Ulm	Bacula ^{10 11} (früher TSM)	Server- und Arbeitsrechner

¹ Info zu Freiburg:

<http://www.rz.uni-freiburg.de/services/sicherheit/tsm> [letzter Zugriff 30.6.2015]

² TSM Tivoli Storage Manager: <http://www-03.ibm.com/software/products/de/tivostormana> [letzter Zugriff 30.6.2015]

³ Info zu Heidelberg:

<http://www.urz.uni-heidelberg.de/service-katalog/serverdienste/backup/> [letzter Zugriff 30.6.2015]

⁴ Info zu Hohenheim:

<https://kim.uni-hohenheim.de/95431> [letzter Zugriff 30.6.2015]

⁵ Info zu Karlsruhe (KIT):

<http://www.scc.kit.edu/dienste/7910.php> [letzter Zugriff 30.6.2015]

⁶ Info zu Konstanz:

<https://www.rz.uni-konstanz.de/dienste/backup/> [letzter Zugriff 30.6.2015]

⁷ Info zu Mannheim:

http://www.uni-mannheim.de/rum/ueber_uns/arbeitsgruppen/zs/zentraler_backup_dienst/ [letzter Zugriff 30.6.2015]

⁸ Info zu Stuttgart:

http://www.tik.uni-stuttgart.de/dienste/datensicherung/backup/Datensicherung_mit_TSM.html [letzter Zugriff 30.6.2015]

⁹ Info zu Tübingen:

<http://www.uni-tuebingen.de/en/facilities/zentrum-fuer-datenverarbeitung/dienstleistungen/storage/zugang-und-kosten.html#c4740> [letzter Zugriff 30.6.2015]

¹⁰ Info zu Ulm:

<http://www.uni-ulm.de/einrichtungen/kiz/it/datenhaltung/backup-service.html> [letzter Zugriff 30.6.2015]

¹¹ Bacula: <http://www.baculasystems.com/> [letzter Zugriff 30.6.2015]

Tab. 11: Überblick der Backup-Dienste.

6.1.4 Handlungsempfehlungen

6.1.4.1 Bereitstellung ausreichender zentraler Speicherkapazitäten

Zurzeit sind die Speicherkapazitäten, die den Forschenden von den Rechen- und Informationszentren zur Verfügung gestellt werden, relativ eingeschränkt. Oft müssen Forschungseinrichtungen ihre Fileserver selbst einrichten, um genügend Speicherplatz zu besitzen. Denkbar wäre die Organisation eines skalierbaren Speicherangebots an den Universitäten, welches von den Rechen- und Informationszentren eingerichtet wird. Dabei könnte man am Anfang des Jahres an alle Wissenschaftsbereiche der Universität ein Angebot für Speicherplatz mit dem Preis pro GB pro Jahr versenden. Forschungseinrichtungen könnten eine Bestellung an das Rechen- und Informationszentrum abgeben. Das Rechen- und Informationszentrum erweitert entsprechend diesen Bestellungen die Speicherkapazitäten im Laufe von mehreren Monaten. Wenn diese Vorgehensweise einmal etabliert ist, wäre es eventuell möglich, bestimmte Vorhersagen für den Speicherbedarf zu treffen. In diesem Fall würde das Rechen- und Informationszentrum einen gewissen Speichervorrat beschaffen, welcher für Forschungseinrichtungen bzw. für Projekte bei Bedarf schnell zur Verfügung gestellt werden kann. Ähnliche Vorgehensweise wurden bereits an einigen Landesuniversitäten etabliert. Dabei betragen bspw. in Tübingen die Kosten für die Forschenden 1,50 Euro pro Monat pro 50 GB Speicherplatz. Die Einrichtung eines zentralen Landesspeicherservice bzw. Erweiterung und Optimierung von bwSync&Share wäre auch denkbar.

6.1.4.2 Bereitstellung zentraler Datenspeicherdienste für personenbezogene Daten

Die aktuellen zentralen Speicherdienste sind zur Speicherung von personenbezogenen Daten nicht geeignet, sodass die Nutzung solcher Dienste für bestimmte Forschungsgruppen nicht möglich ist. Es wäre daher sinnvoll, effiziente technische und organisatorische Verfahren zu etablieren, um auch die Speicherung von personenbezogenen Daten zu ermöglichen (s. [2.1. Urheberrecht und Datenschutz](#)).

6.1.4.3 Ausbau der Bandbreite der Datenleitungen

Da die Datenmenge kontinuierlich steigt, steigt auch der Bedarf nach schnelleren Datenleitungen. Deshalb ist die Fortsetzung eines konsequenten Ausbaus der Datenleitungsbandbreite in der Forschungsinfrastruktur in Baden-Württemberg¹⁷⁴ notwendig.

6.1.4.4 Erhöhung der Benutzerfreundlichkeit von IT-Diensten

Die Erfahrung aus dem Projekt zeigt, dass Landesdatendienste, deren Nutzung umständlich und zeitaufwändig ist, oder die insgesamt instabil sind, nicht genutzt werden, sondern hier auf kostenlose frei zugängliche kommerzielle Dienste, wie beispielsweise DropBox, zurückgegriffen wird, selbst wenn deren Nutzung datenschutzrechtlich bedenklich ist. Deshalb ist eine weitere Optimierung von bereits existierenden Diensten wie bwSync&Share, bwHPC, Backup

¹⁷⁴ BelWü - das Landeshochschulnetz <https://www.belwue.de/> [letzter Zugriff 30.6.2015].

(TSM) usw. notwendig. Für die Messung von Benutzerfreundlichkeit wäre eine Auswertung von anonymisierten Nutzungsstatistiken sowie die Durchführung von Usabilitytests sinnvoll. Als Kriterien für Benutzerfreundlichkeit könnten dabei u. a. steigende Nutzerzahlen mit der gleichzeitigen Senkung der Anzahl von Supportanfragen bei Problemen genutzt werden.

6.1.4.5 Beratung zur IT-Infrastruktur

Nicht alle IT-Angebote, die alltägliche Forschungsarbeit einfacher und effizienter machen könnten, sind Forschenden bekannt. Deshalb stellt die Erweiterung von Beratungsmöglichkeiten eine sinnvolle Ergänzung der Weiterentwicklung von Forschungsinfrastruktur in Baden-Württemberg dar (s. [2.2. Informationsangebote](#)).

6.1.4.6 Integration verschiedener IT-Diensten in virtuelle Forschungsumgebungen

Wenn sich Forschende keine Gedanken mehr machen müssen, wo sie ihre wachsenden Datenmengen speichern können, wie sie gemeinsam mit Projektpartnern ihre Projektdaten sowie gemeinsame Schriften einfach in einem System bearbeiten können, ist das ein großer Schritt zur Etablierung virtueller Forschungsumgebungen. Teilweise ist diese Entwicklung bereits eingeleitet, beispielsweise durch die Verwendung gleicher Login-Daten bei Nutzung verschiedener Landesdienste im Rahmen des bwIDM-Projektes. Ebenso repräsentiert die Nutzung des bwSync&Share-Cloudspeichers die Entwicklung in die Richtung der virtuellen Forschungsumgebungen. Diese Entwicklung sollte nun konsequent fortgesetzt werden. U. a. scheint es sinnvoll, den Forschenden folgende IT-Werkzeuge zur Verfügung zu stellen:

- Tools zum kollaborativen Arbeiten
- Archivierungsdienste, um Forschungsdaten langfristig archivieren zu können
- einfache Einrichtung virtueller Maschinen, um gemeinsam gleiche Anwendungen, unabhängig vom Betriebssystem auf dem eigenen Arbeitsrechner, zu nutzen.

6.2 IT-Support (219)

6.2.1 Einleitung

IT-Support ist ein essenzieller Teil zur effizienten und effektiven Nutzung der immer komplexeren IT-Infrastruktur. Schnelle Hilfe bei EDV-Problemen erlaubt die Konzentration auf die eigentliche Forschungsarbeit statt wertvolle Zeit in die Lösung von EDV-Störungen zu investieren. Im aktuellen Kapitel wurden die in Interviews erfassten Wünsche von Forschenden im Bereich des IT-Supports analysiert.

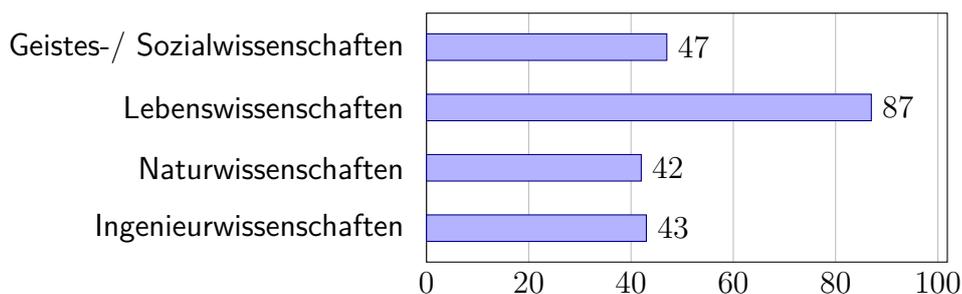


Abb. 42: Verteilung der User-Stories nach der Fachsystematik der DFG

6.2.2 Bedarf der Wissenschaftsgemeinschaft

„Als Forschende, wünschen wir uns mehr IT-Support, um durch die IT-Infrastruktur in der eigenen Forschung effizient unterstützt zu sein.“

Dem Bereich IT-Support wurden 219 User-Stories zugewiesen. Eine Vielzahl davon stammen von den Forschenden aus dem Bereich Lebenswissenschaften (Abb. 42).

Insgesamt ging es den Forschenden beim Ausbau des IT-Supports darum, mehr Zeit in die Forschung investieren zu können und weniger Zeit bei der Lösung von IT-Problemen verschwenden zu müssen. Das Spektrum umfasst dabei die Hilfe beim Einrichten sowie dem Betreiben von Diensten, Servern und Arbeitsplatzrechnern. Bedarf nach Bereitstellung von komplett eingerichteten Arbeitsrechnern wurde von mehreren Forschenden geäußert. Auch schnellerer Support bei IT-Problemen wird gewünscht. Ausbau von Unterstützung für Linux- und Apple-Systeme sowie für Open-Source-Anwendungen wurde ebenso artikuliert. Der Aspekt der Supportnachhaltigkeit war ebenso Gegenstand der ausgesprochenen Bedarfe. Beispielsweise werden viele Messgeräte in Kombination mit ihren Steuerrechnern in Forschungseinrichtungen Jahre, manche sogar über Jahrzehnte, genutzt. Der Wunsch nach einem langfristigen Support für solche Systeme wurde von Forschenden als erstrebenswert angesehen. Ferner ging es den Forschenden um Vereinheitlichung und Erhöhung der Verfügbarkeit der Basis-IT-Dienste. Um den Ausbau des IT-Supports an den Universitäten zu fördern, wurde u. a. eine Erweiterung der Finanzierung des IT-Bereichs der Universitäten für sinnvoll befunden.

Zum Thema „Beratungen, Schulungen und Informationswege“ wurden zusätzliche Bedarfe ausgesprochen. Dabei wurde u. a. eine bessere Abstimmung zwischen den IT-Beauftragten der Institute und den Rechen- und Informationszentren gewünscht. Auch Bedarfe über Klarheit der Ansprechpersonen bei IT-Fragen und ein besserer Informationsfluss wurden artikuliert. Als mögliche Informationswege wurden beispielsweise Alert-Mails und Newsfeeds mit Informationen über neue IT-Dienste sowie Weiterbildungsangebote genannt.

6.2.3 Status quo

An allen Landesuniversitäten übernehmen Rechen- und Informationszentren die Funktion des EDV-Supports. Es gibt an jeder Universität Hotline-Services als Hilfestellung bei Problemen (Tab. 12). Die Erreichbarkeit ist zwischen 8:00 bis 15:00-18:00 abhängig von der jeweiligen Universität und ist während der Arbeitswoche gegeben. Außerdem können Forschende direkt

bei einem entsprechenden IT-Service-Desk persönlich vorbeikommen und ihre Rechner bei Problemen zur Überprüfung und Reparatur abgeben sowie notwendige Beratung bekommen. Online-Support per Teamviewer ist in Freiburg möglich.

Standort	Dienste	Link [letzter Zugriff 17.6.2015]
Freiburg	Online HelpCenter, Ansprechstellen, Online Support per Teamviewer	http://www.rz.uni-freiburg.de/helpcenter
Heidelberg	IT-Service-Support, IT-Service-Point, IT-Service-Hotline	https://it-service.uni-heidelberg.de/
Hohenheim	IT-Service-Desk mit Support-Hotline	https://kim.uni-hohenheim.de/bedienstete-startseite#jfmulticontent_c221501-3
Karlsruhe	IT-Service-Desk mit Support-Hotline	https://www.scc.kit.edu/hotline/index.php
Konstanz	Support mit Hotline	https://www.rz.uni-konstanz.de/support/
Mannheim	Support mit Hotline	http://www.uni-mannheim.de/rum/support/
Stuttgart	Support mit Hotline	http://www.tik.uni-stuttgart.de/kontakt/support.html
Tübingen	Hotline / Beratung	http://www.uni-tuebingen.de/en/facilities/zentrum-fuer-datenverarbeitung/kontakt-antraege-beratung/hotline.html
Ulm	Helpdesk, Hotline, Kundenportal	http://www.uni-ulm.de/einrichtungen/kiz/it.html

Tab. 12: IT-Support an den Universitäten in Baden-Württemberg

Auf den Webseiten von IT-Abteilungen der Universitäten gibt es außerdem Informationen zu IT-Diensten sowie Anleitungen zu ihrer Nutzung. Support-Aufträge können mit Hilfe von Webformularen erstellt und an ein Ticketsystem weitergeleitet werden.

Es gibt aber nicht genügend Ressourcen in Rechen- und Informationszentren, um umfassenden IT-Support für die Forschungseinrichtungen zu erreichen. Ein Teil der Aufgaben zum EDV-Support wird von IT-Beauftragten bzw. von IT-affinen Forschenden der jeweiligen Universitätseinrichtungen erfüllt. Teilweise richten Forschende ihre Arbeitsplatzrechner selbst ein. Das betrifft zum Teil auch das Einrichten und Betreiben von Arbeitsgruppenservern, was nicht zuletzt den individuellen Bedürfnissen geschuldet ist.

6.2.4 Handlungsempfehlungen

6.2.4.1 Ausbau des EDV-Supports, inkl. mehr Personal

Es wurde oft der Wunsch ausgesprochen, dass Forschende weniger Zeit in Lösung von EDV-Problemen investieren möchten, um sich mehr auf ihre Forschungsarbeit konzentrieren zu können. Deshalb wäre die Erweiterung des IT-Supports sinnvoll. Eine Möglichkeit bestünde in dem Ausbau der bereits existierenden Angebote und wartungsfreundlicher Systeme. Die versteckten Kosten der „EDV-Betreuung in Eigenregie“ sollten deutlicher in die bisherigen EDV-Verwaltungskonzepte eingerechnet werden. Dabei ist zu bemerken, dass einige Disziplinen rein zentralen Rootzugriff auf ihre Rechner extrem kritisch sehen.

6.2.4.2 Verbesserung des Informationsflusses

Um den Informationsfluss zwischen den Rechen- und Informationszentren und Forschenden zu verbessern, könnte der Ausbau des Beratungsangebots, sowie Optimierung der Informationswege in Frage kommen (s. 2.2. [Informationsangebote](#)). So wurden an einigen Standorten beispielsweise Formate zur Kommunikation zwischen den IT-Beauftragten der Institute und den zentralen IT-Einrichtungen erprobt. Für andere Universitätsmitglieder könnten ähnliche Informations- und Beratungsangebote, sowie Weiterbildungskurse sinnvoll sein, um ihren Umgang mit IT-Problemen und IT-Fragestellungen effizienter gestalten zu können.

6.3 Lizenzierung (167)

6.3.1 Einleitung

Die gewonnenen User-Stories der *Lizenzierung* thematisieren hauptsächlich die Beschaffung von Softwarelizenzen oder den Zugriff auf Journale bzw. Repositorien oder Datenbanken. Der Softwareaspekt (von Analysetools) ist bei den Lebenswissenschaften besonders ausgeprägt, weshalb sie aus dem Gesamtbedarf etwas hervorstechen.

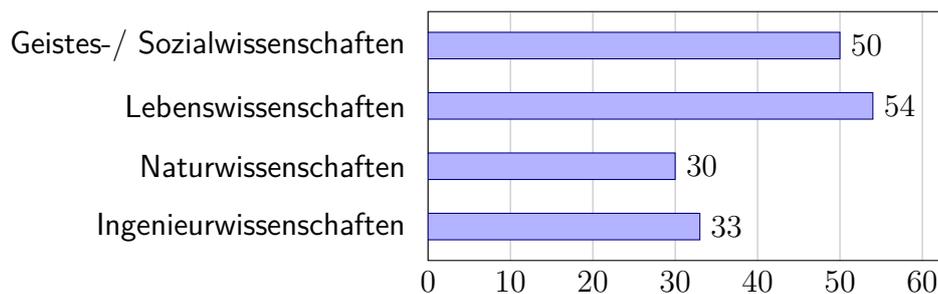


Abb. 43: Verteilung der User-Stories nach der Fachsystematik der DFG

6.3.2 Bedarf der Wissenschaftsgemeinschaft

„Als Forschende, die eine zentrale, günstige Lizenzlösung bei der Beschaffung von Software für effektiv halten, wünschen wir uns generell mehr Campuslizenzen für Software, um Software günstiger und einfacher beschaffen zu können.“

„Als Mitglieder einer interdisziplinären Forschungsgruppe wünschen wir uns Lizenzen, um Zugang zu den Journalen [und / oder Repositorien] (außerhalb) unseres Fachgebiets zu haben.“

Werden kommerziell angebotene Software oder Journale für Forschungstätigkeiten benötigt, besteht der Wunsch nach einer effizienten und kostengünstigen Zugangsmöglichkeit zu den Angeboten.

Außer diesen beiden Hauptrichtungen bestehen Unsicherheiten bei den Forschenden bzgl. lizenzierter Angebote wegen Grauzonen im Rechtsbereich. Ein unklarer Aspekt ist wie es unter verschiedenen Gesichtspunkten einzuordnen ist, wenn ein Programm Internetzugang für die Benutzung erfordern. Eine weitere Unklarheit ist die Insolvenz von Softwareherstellern. Forschende sind sich unsicher, ob sie Software von kleinen und unbekanntem Unternehmen benutzen sollen, weil sie nicht wissen, wie verlässlich diese in Zukunft sein werden. Manche Forschungstätigkeiten beruhen auf langjährigen Studien und die Softwareprodukte müssen für diese entsprechend gepflegt werden, um dauerhaft und gleichartig verfügbar zu sein. Wenn ein Unternehmen jedoch proprietäre Softwareschnittstellen benutzt kann dies zu Schwierigkeiten führen, falls es beschließt gewisse Softwareeigenschaften zu entfernen oder sich insolvent meldet.

6.3.3 Status quo

Einzelne Institute können die Kosten für manche Softwareprodukte nicht für alle Mitglieder decken und benutzen deswegen Einzelplatzlizenzen. Diese Tatsache hindert die Forschungstätigkeit insbesondere dann, wenn verschiedene Bedingungen an die Benutzung der Software gebunden sind. In der Summe fallen ggf. hierdurch höhere Kosten als bei einer vom Land beauftragten Volumenbeschaffung an. Zurzeit existiert noch keine einheitliche Vorgehensweise bei der Beschaffung von Software¹⁷⁵. Die einzelnen Universitäten sind für die Beschaffungen selbst verantwortlich und tragen die Kosten eigenständig. Eine gemeinsame Koordination dieser Beschaffungen wäre sinnvoll und anzustreben.

6.3.4 E-Science-Bestrebungen

Das Thema der Lizenzierung ist eines der vier großen strategischen Konzepte, welche im ESFK behandelt werden. Die Wichtigkeit der Thematik wurde bereits erkannt und erste Ziele definiert¹⁷⁶:

¹⁷⁵ Die einzige in Landeskoooperation stattfindende Beschaffung ist die bwNotebook-Initiative, welches von der Universität Freiburg verwaltet wird.

¹⁷⁶ ESFK, 12.

- „Zum Studieren, Lehren und Forschen in Baden-Württemberg werden die benötigte Literatur bzw. andere Informationsmedien unmittelbar zur Verfügung stehen, sei es Open Access, mit vorhandener Lizenz oder als elektronischer Sofortlieferdienst.“
- „Die Informationsversorgung mit elektronischen Fachinformationsangeboten für Forschung und Lehre wird auf drei Ebenen organisiert: über lokal beschaffte und vorgehaltene Lizenzen, konsortial erworbene Lizenzen und durch die DFG geförderte National- bzw. Allianz-Lizenzen.“
- „Das Konsortium wird gegenüber den Verlagen und Anbietern von der Universitätsbibliothek Freiburg vertreten.“

6.3.5 Gegenwärtige Handlungsempfehlungen

6.3.5.1 Workflow für die Beantragung von Landeslizenzen

Es sollte ein Workflow für die gemeinsame Lizenzbeschaffung entworfen werden. Die Schwierigkeiten hierbei liegen in der Kostenverteilung und Benutzerübersicht, welche ggf. an das Softwareunternehmen weitergegeben werden muss. Für die Ermittlung des Bedarfs für eine Software kann eine Webseite benutzt werden. Die Forschenden könnten dort Vorschläge für die Beschaffung angeben oder existierende Vorschläge anderer unterstützen. Weiterhin könnte eine Hilfe / Diskussion zu Alternativsoftware geführt werden. Diese Initiative, verknüpft mit der Unterstützung von Open-Source-Software, könnte als zentraler Einstieg für Empfehlungen benutzt werden, um Forschende über Open-Source-Alternativen für bestimmte Programme zu informieren. Die Webseite kann zusätzlich weitere Experten zu den Programmen einbinden, welche bei Problemen und Fragen helfen können. Wichtige Eigenschaften einer solchen Webplattform sind:

- Empfehlung von Programmen zur Beschaffung als Landeslizenz
- Abstimmung über die zu beschaffenden Programme
- Möglichkeit zur Bildung von Anwender-Communitys zum Austausch von Best-Practice-Lösungen für die jeweilige Software
- Empfehlungen für Open-Source-Programme, anstatt der gewünschten kommerziellen Lösung

6.3.5.2 Benutzung im ungeklaren Rechtsraum

Wie bereits beschrieben existieren verschiedene Grauzonen in der Lizenzbeschaffung bzw. Benutzung von Software in der Wissenschaft. Es ist zu klären, was mit der Software passiert, falls der Hersteller nicht mehr existiert oder die weitere Benutzung von Lizenzen aufgrund der Abschaffung von Softwarefeatures nicht mehr zielführend ist. Welche Rechte und Pflichten sollten die Forschenden im Vorfeld vom ehemaligen Softwarehersteller verlangen (insbesondere im Bereich Langzeitsupport), um auch weiterhin die Software in ihrer Forschung zu

benutzen? Durch die Bildung eines Arbeitskreises können diese Fragen beantwortet und den Forschenden durch Empfehlungen und Handreichungen eine gewisse Rechtsklarheit gegeben werden.

6.4 Finanzierung (67)

6.4.1 Einleitung

Dieses Kapitel behandelt den Themenbereich Finanzierung und solche User-Storys, die in der Hauptsache einen Zustand der Finanzierung zu ändern fordern und noch nicht ausreichend von anderen Kapiteln abgedeckt sind. So beinhalten bspw. Wünsche zur Lizenzierung, zur Unterstützung der Metadatenanreicherung oder zur Pflege der Forschungsinfrastruktur der Gruppe durch Schlüsselpersonal (z. B. FDM-Beauftragter) oft einen starken strukturellen Finanzierungswunsch, sind aber in ihren jeweiligen Kapiteln schon einmal behandelt.

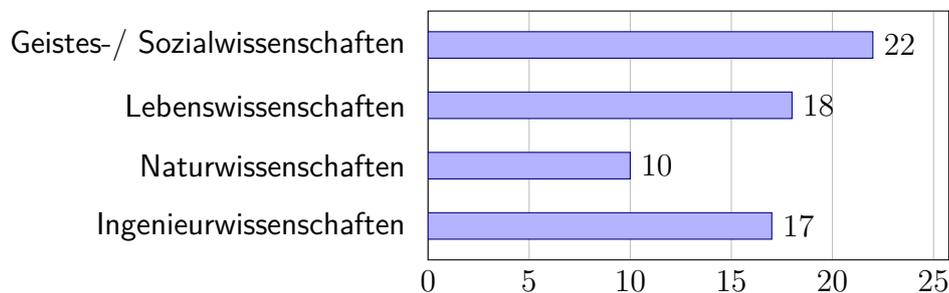


Abb. 44: Verteilung der User-Storys nach der Fachsystematik der DFG

6.4.2 Bedarf der Wissenschaftsgemeinschaft

Es verbleiben 67 User-Storys die gemäß [Abb. 44](#) auf alle Wissenschaftsbereiche verteilt sind.

6.4.2.1 Kontinuität (23)

„Als Forschende, wünschen wir uns Kontinuität in der Finanzierung, um langfristig besser forschen zu können.“

Von diesen 23 User-Storys hat die Mehrheit vor allem ihr Personal und die Probleme des Wissensabflusses im Fokus, die sich durch hohe Personalfuktuation ergeben, der vor allem bei der Besetzung von Schlüsselpositionen nur schwer zu verkraften ist. Drei User-Storys aus den Ingenieurwissenschaften beklagen in diesem Zusammenhang auch, dass es unmöglich sei hochkompetentes (IT-)Personal durch das Schaffen attraktiver Stellen (entfristet, flexibel, gut bezahlt) zu halten, um eine Abwerbung aus der Industrie zu verhindern. Der Fokus der anderen fünf User-Storys geht eher in eine strukturelle Richtung und beklagt die Erschwernis langfristig ausgerichteter Forschung durch eine ständige „Projektkultur“.

6.4.2.2 Strukturelles (27)

„Als Forschende, wünschen wir uns Änderungen an der aktuellen Finanzierungsstruktur, damit Mittel dahin fließen, wo sie optimal eingesetzt werden können oder zumindest mehr Klarheit und Fairness bei der Verteilung herrscht.“

Einige Forschende, wünschen sich mehr Flexibilität, wenn beispielsweise in einem Projekt (plötzlich) eine zusätzliche Anschaffung (oft Hardware oder Software) sinnvoll scheint, diese aber weder aus den Projektmitteln noch aus den Grundmitteln bezahlt werden kann. Sechs Forschende machen sich in dem Zusammenhang Gedanken zur Modernisierung von Hardware oder Geräten, die auf der einen Seite nicht ständig durchgeführt werden kann, auf der anderen Seite aber ständig gefordert ist, um Spitzenforschung zu betreiben. Andere wünschen sich mehr Klarheit oder Regeln, wie universitätsinterne Dienstleistungen abrechenbar sind (u. a. Datenanalyse) oder grundsätzlich eine andere Behandlung von Overheadkosten. Für wieder andere steht der faire Wettbewerb im Fokus oder dass sich aus öffentlichen Geldern subventionierte Industriepartner auch an den Zweck der Subventionsgewährung halten.

6.4.2.3 Langfristige Ergebnissicherung (13)

„Als Forschende mit hochwertigen Projektergebnissen, wünschen wir uns eine Lösung, um unsere Ergebnisse langfristig nach Projektende (ohne weitere Betreuung) in sinnvoller Weise zu sichern.“

Jeder der hier aufgeführten User-Stories ist nicht einfach mit einem Standard-Repositoryum befriedigt. Beispielsweise kann die Darstellung der Forschungsergebnisse einen Server oder viel Speicherplatz benötigen. Sechs der User-Stories lassen sich auf den gemeinsamen Nenner bringen, dass das „kleine selbst aufgebaute Repositoryum“ an anderer Stelle weiter gepflegt werden soll.

6.4.2.4 Fachspezifisches (6)

„Als Forschende, wünschen wir uns, dass gewisse Aspekte unserer Disziplin besser von den Forschungsförderern berücksichtigt werden, um Förderung für für uns wichtige Anliegen erhalten zu können.“

Die sechs heterogenen Wünsche sind:

... eine bessere finanzielle Ausstattung der Gruppe für die Erwerbung von exotischen Substraten und Intermediaten, um diese oft extrem teuren, aber für die Forschung wichtigen Sachmittel erwerben zu können.

... mehr finanzielle Ausstattung durch die Universität, um auch riskante Projekte zu finanzieren, für die es schwierig ist, Drittmittel zu bekommen.

... Unterstützung forschender Mediziner vom Land und Bund, um mehr Mediziner in die Forschung zu bewegen und somit die Modelle und Verfahren zu verbessern, weil aktuell die Forschung eine Tätigkeit darstellt, welche die Mediziner in ihrer Freizeit tun und nicht entschädigt werden.

... Finanzierung für die Vergabe von Aufträgen zur Ganz-Genom-Sequenzierungen, um nicht nur mit partiellen Sequenzen arbeiten zu müssen.

... , dass unsere Mischkultur aus Print- und Onlinemedien sowie deren Methoden berücksichtigt werden, um zu vermeiden, dass beispielsweise Druckkostenzuschüsse der DFG wegfallen. ... , dass bei der Mittelbewilligung für Veröffentlichungen (insbesondere Bücher) auch PR-Aspekte einbezogen werden, um die Veröffentlichungskosten von der Universität leichter gedeckt zu bekommen.

6.4.3 Status quo

Eine Erörterung des Status quo der Finanzierung der Forschungslandschaft findet hier natürlich nicht vollständig statt. Das BMBF beschreibt umfassend die Aufgaben und Bereiche der DFG¹⁷⁷. Bis 2014 durfte die DFG die Hochschulen aufgrund des Kooperationsverbots zwischen Bund und Ländern nur in thematisch begrenzten Projekten fördern. Trotz der zunächst neuen Gesetzeslage sind grundsätzliche Änderungen der Finanzierungspolitik der DFG kurzfristig nicht zu erwarten. Außerdem gibt es eine (un)klare¹⁷⁸ Trennung zwischen der Grundfinanzierung, die grundsätzlich von der Universität (bzw. deren Bundesland) geleistet werden soll (Büro, Grundausstattung usw.) und der Projekt(-Zusatz-)förderung der DFG und anderer Förderorganisationen. Hier werden je nach Förderer auch prozentuale Overheadmittel zur Verfügung gestellt, die die allgemeinen Mehrausgaben an den Universitäten durch die Projekte ausgleichen sollen.

6.4.4 Handlungsempfehlungen

Grundsätzlich wäre es wünschenswert, wenn für die Langfristigkeit und Kontinuität von Projektergebnissen über das jeweilige Projektende hinaus mehr Mittel zur Verfügung stünden. Von den Fördermittelgebern sollte dazu überlegt werden, wie man geschickt „die zweitbeste Forschung, die viele nachnutzen können“ höher gewichtet als „das beste Forschungskonzept im eigenen Keller“. Wir empfehlen dazu dringend eine entsprechende Leitlinie und die flächendeckende Einbeziehung von Datenmanagementplänen in die Begutachtung, selbst wenn diese in einer Disziplin nicht nötig oder üblich sind. Das sollten die Gutachter bewerten. Weitergehende Handlungsempfehlungen für das Kapitel Finanzierung verschriftlichen wir ausdrücklich nicht, weil dieses Thema zwar sehr wichtig ist, aber zu weit entfernt vom Projektfokus von bwFDM-Communities liegt.

6.5 Open Science (89)

6.5.1 Einleitung

Zurzeit existieren verschiedene europa-, bundes- und landesweite Ansätze, mehr Open-Access-Möglichkeiten in der Wissenschaft zu etablieren. Open Access hat das Ziel „*wissenschaftliche Literatur und wissenschaftliche Materialien für alle Nutzerinnen und Nutzer frei zugänglich zu*

¹⁷⁷ <http://www.bmbf.de/de/251.php> - [Letzter Zugriff am 30.6.2015].

¹⁷⁸ Forschende hatten sich beschwert, dass sie bei Finanzierungsgesuchen bspw. zu Hardware-Erweiterungen zwischen Uni, Land und DFG im Kreis herum geschickt wurden.

machen: kostenlos und möglichst frei von technischen und rechtlichen Barrieren. Ein wesentlicher Aspekt ist die Maximierung der Verbreitung wissenschaftlicher Information.“¹⁷⁹. Open Access ist aber nur der erste Schritt auf dem Weg zur offenen Wissenschaft. Denn erst wenn Ergebnisse präzise reproduzierbar und beispielsweise für die Gesellschaft ohne Zusatzsoftware visualisierbar sind, ist die Wissenschaft an sich (und nicht nur abgeleitete Texte) tatsächlich offen.

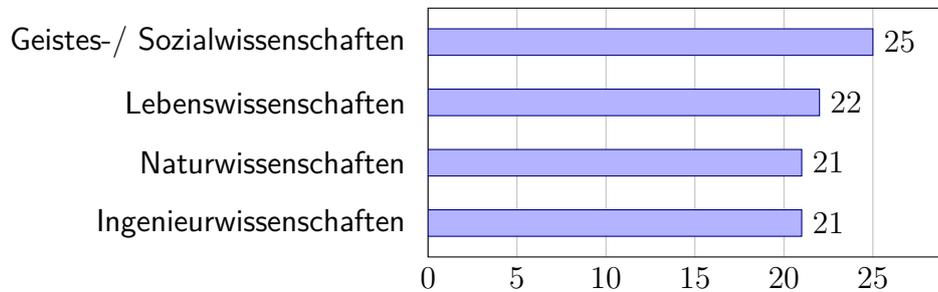


Abb. 45: Verteilung der User-Stories nach der Fachsystematik der DFG

6.5.2 Bedarf der Wissenschaftsgemeinschaft

Es ergibt sich natürlicherweise ein thematischer Überschneidung zu 2.3. *Wissenschaftliche Datenkultur* sowie 5.2. *Repositorien für die Publikation von Forschungsdaten* und 3.3. *Formate und einheitliche Standards*. Etwa 40 % der hier beschriebenen User-Stories haben Berührungspunkte mit einem dieser drei Themenfelder. Nichtsdestotrotz fallen rund drei Viertel der Open-Science-Stories unter zwei wichtige, bisher noch nicht behandelte Aspekte. Diese sind Open Source und Open Access (letzteres jenseits von Daten und Software, meist mit Bezug zu Journalen). Die restlichen User-Stories sind eine Mischung verschiedener Wünsche nach besserer wissenschaftlicher Praxis oder Möglichkeiten dafür. Dabei stehen oft datenferne Aspekte im Mittelpunkt, die bisher noch nirgends angesprochen sind.

6.5.2.1 Open Access (41)

„Als Publizierende, wünschen wir uns eine bessere Unterstützung von Open Access, um wissenschaftliche Erkenntnisse ohne Einschränkungen verbreiten zu können.“

Viele Forschende haben den Wunsch ausgedrückt, dass das Land die Kosten der Veröffentlichung in Open-Access-Journalen übernimmt. Weiterhin wurde von vielen Forschenden das Fehlen einer Infrastruktur zu freien Veröffentlichung von wissenschaftlichen Ergebnissen bemängelt. Der Wunsch hierbei ist, die Kosten und Verwaltung bei der Veröffentlichung von Daten auf das Land zu übertragen, sodass sich die Forschenden auf die Forschung konzentrieren können. Auch dass Journale durch Restriktionen, langsame Bearbeitungszeiten und einen falschen Fokus (Verkaufszahlen) dem Zweck einer effizienten Wissensvermittlung häufig nicht dienlich werden, wurde mehrfach bemängelt.

¹⁷⁹ <http://www.openaccess-germany.de/informationen-zu-open-access/was-bedeutet-open-access/> [letzter Zugriff 14.5.2015].

Forschende investieren sehr viel Arbeit in die Erhebung und Analyse ihrer Daten. Deshalb ist es nur verständlich, dass sie ohne Anreiz anderen ungern ihre Daten überlassen. Gewisse Zusicherungen von Karenzzeiten, um die Daten zumindest für eine bestimmte Zeit vor fremden Augen zu schützen oder die Hoheit der Verwertung beim Produzenten zu lassen ist ein Punkt, dessen Spannung zu Open Access abgebaut werden soll. Die geforderte Karenzzeit soll dazu verwendet werden, die Daten entsprechend zu analysieren und ggf. aufzubereiten.

6.5.2.2 Open Source (26)

„Als Forschende die ungern proprietäre Software nutzen, wünschen wir uns mehr freie Programme und einheitliche Standards sowie einen langlebigeren Support für Open Source, um freien Austausch von Informationen zu ermöglichen.“

Über ein Viertel der Wünsche im Bereich Open Access beziehen sich auf Open-Source-Software. Es wird besserer Support, mehr Funktionalitäten und der Einsatz von freien Formaten gewünscht. Ein Unabhängigkeitswunsch von proprietärer Software wie MatLab, Origin oder Ansys und die Entwicklung von Alternativen zu diesen Programmen, wurde mehrmals ausgedrückt. Die Forschenden legen großen Wert auf die Benutzung bzw. Unterstützung von Open-Source-Software. Es existieren verschiedene Bereiche, in denen der Einsatz von Open-Source-Software zur Zeit nicht möglich ist. Die Forschenden fordern auch in diesen Bereich die Festlegung von Standards und Entwicklung von Open-Source-Software. Es tauchte auch der Aspekt auf, dass Open Source allein nicht reicht, wenn sich niemand um die langfristige Funktionalität der Software kümmert.

6.5.2.3 Würdigung hoher Forschungsqualität (25)

„Als Forschende, wünschen wir uns einen stärkeren Fokus auf Forschungsqualität und Angebote dafür, um das was die Wissenschaft voranbringt auch entsprechend zu würdigen und zu unterstützen.“

Um die Qualität der Aufbereitung sicherzustellen, sind verschiedene Reviewverfahren vorgeschlagen worden, um über die Wiederverwendbarkeit zu entscheiden. Dabei kamen mehrfach auch Ideen zur Ausgestaltung eines offenen Reviews auf. Aber auch Tools, um beispielsweise Forschung miteinander besser zu vergleichen sowie ein anderer Umgang mit wissenschaftlichen Erkenntnissen bei Medien, Gerichten und Journalen wird gefordert. Viele meinen, dass es Zeit wäre für eine Grundsatzdiskussion darüber, wie man Forschung evaluiert und die bisherigen Verfahren neu gestaltet.

6.5.3 Status quo

Alle Universitäten in Baden-Württemberg unterstützen die „Berliner Erklärung“ zu Open Access¹⁸⁰. Im Gegensatz zu den klassischen Verlagen haben Open-Access-Zeitschriften ein anderes Finanzierungsmodell. Die Autoren der einzelnen Beiträge bezahlen die Veröffentlichungen selbst und die Leser bekommen freien Zugang zu den Zeitschriften. Aus diesem

¹⁸⁰ <http://openaccess.mpg.de/Berliner-Erklaerung> [letzter Zugriff 14.5.2015].

Grund haben viele Universitäten einen Publikationsfond in Kooperation mit der DFG eingerichtet, mit dessen Hilfe Open-Access-Veröffentlichungen abgerechnet werden können. Wir kennen die Vereinbarungen natürlich nicht alle, erwarten aber, dass die DFG langsam verlangt, dass die Universitäten diesen Fond selbst stemmen.

Aktuelle Unterstützungsmöglichkeiten von Open Science:

- Publikationsserver für die freie zugängliche Veröffentlichung von wissenschaftlichen Artikeln
- Repositorien für die Bereitstellung von Forschungsdaten für andere Forschende
- Journalsoftware für die Erstellung und Veröffentlichung eines eigenen Open-Access-Journals
- Informationsplattform über aktuelle Forschungsprojekte an der Universität (s. 2.2. Informationsangebote)

Im Folgenden werden die aktuellen Förderungen von Open Access an den einzelnen Universitäten in Baden-Württemberg vorgestellt.

Karlsruhe Forschenden am Karlsruher Institut für Technologie haben verschiedene Dienstleistungen, unter denen sie auswählen können. Das KIT bietet zwei Dokumentenserver an (EVA STAR¹⁸¹ und KAROLA¹⁸², welche die Forschenden für Primärveröffentlichungen oder Parallelveröffentlichungen benutzen können. Weiter existiert der Verlag „KIT Scientific Publishing“¹⁸³, welcher ebenfalls für die Veröffentlichung von Arbeiten im Internet oder im Buchhandel genutzt werden kann. Neben der Kostendeckung von Publikationen durch den KIT-Publikationsfond¹⁸⁴ können Forschende vom KIT aufgrund von institutionellen Mitgliedschaften die Publikationskosten in Open-Access-Zeitschriftenportalen BiomedCentral (mit ChemistryCentral und SpringerOpen), Copernicus Publications, Frontiers, MDPI, PLOS und Wiley Open Access erstattet werden.

Freiburg Die Universität Freiburg bietet einen eigenen Dokumentenserver FreiDok¹⁸⁵. Auf FreiDok können Dissertationen, Aufsätze, Proceedings oder andere wissenschaftliche Artikel zur freien Verfügung gestellt werden. Zudem fallen für Forschende an der Universität Freiburg keine Publikationskosten bei BiomedCentral (mit ChemistryCentral und SpringerOpen) an.

Heidelberg Um die Forschenden beim Einsatz von Open Access zu unterstützen, wurde an der Universität Heidelberg ein Dokumentenserver (heiDOK¹⁸⁶), ein Repository für Forschungsdaten (heiDATA¹⁸⁷), ein Publikationsservice (heiJOURNALS¹⁸⁸) und ein Open-

¹⁸¹ <http://services.bibliothek.kit.edu/eva/index.html> [letzter Zugriff 14.5.2015].

¹⁸² http://opac.fzk.de:81/de/fzk_qsim_frm.html [letzter Zugriff 14.5.2015].

¹⁸³ http://www.ksp.kit.edu/Der_Verlag [letzter Zugriff 14.5.2015].

¹⁸⁴ <http://www.bibliothek.kit.edu/cms/kit-publikationsfonds.php> [letzter Zugriff 14.5.2015].

¹⁸⁵ <https://www.freidok.uni-freiburg.de/> [letzter Zugriff 14.5.2015].

¹⁸⁶ <http://archiv.ub.uni-heidelberg.de/volltextserver/> [letzter Zugriff 14.5.2015].

¹⁸⁷ <https://heidata.uni-heidelberg.de/dvn/> [letzter Zugriff 14.5.2015].

¹⁸⁸ <http://journals.ub.uni-heidelberg.de/index.php/ojs/index> [letzter Zugriff 14.5.2015].

Access-Publikationsfond¹⁸⁹ eingerichtet. Dokumentenserver und Publikationsfond erfüllen die gleichen Aufgaben wie in den vorhergehenden Universitäten auch: Speicherplatz für Artikel und Dissertationen bzw. finanzielle Unterstützung bei der Publikation in Open Access. Der Publikationsservice heiJOURNALS bietet Forschenden die Möglichkeit, eigene Open-Access-Journale zu veröffentlichen. Für diese Dienstleistung kommt die Open-Source-Software *Open Journal Systems* zum Einsatz.

Konstanz In Konstanz wird zusätzlich zu einem Dokumentenserver (KOPS¹⁹⁰), Open-Access-Publikationsfond¹⁹¹ und Publikationssoftware (PKP¹⁹²) die Informationsplattform Sci-Kon¹⁹³ betrieben. Diese Plattform soll eine Übersicht zu den aktuellen Forschungsprojekten an der Universität darstellen. Zusätzlich sind Informationen zu den wissenschaftlichen Mitarbeiterinnen und Mitarbeiter abgelegt und deren Publikationen aus KOPS verlinkt.

Stuttgart Auch in Stuttgart verfügen die Forschenden über einen Open-Access-Publikationsfond¹⁹⁴ und einem Dokumentenserver (OPUS¹⁹⁵) für Parallelveröffentlichungen. Zudem ermuntert die Universität Stuttgart in ihrer Open-Access-Policy¹⁹⁶ vom Februar 2011 die Benutzung dieser Plattformen. Zudem ist die Universität Mitglied bei folgenden Open-Access-Verlagen: BioMed Central, MDPI, Frontiers, PLOS.

Tübingen Tübingen verabschiedete eine Open-Access-Policy im Februar 2013 und bietet ein Publikationsfond¹⁹⁷ in Kooperation mit der DFG an. Zudem wird gerade ein OJS-Angebot¹⁹⁸ aufgebaut, welches Forschenden an in Tübingen helfen soll, eigene Open-Access-Journale zu veröffentlichen. Weiterhin bietet die Universität den Dokumentenserver TOBIAS-lib¹⁹⁹ und TOBIAS-bild²⁰⁰ an. Das Erste dient als Speicherplattform für Dissertationen und andere elektronische Dokumente. Aufgrund der hohen Anzahl an Natur- und Geisteswissenschaften an der Universität wurde der letzte Service TOBIAS-bild eingerichtet. Hier können Forschenden ihre Bilder und Dokumentenabbildungen bzw. Scanbilder speichern und zentral verwalten.

¹⁸⁹ <http://www.ub.uni-heidelberg.de/service/openaccess/publikationsfonds.html> [letzter Zugriff 14.5.2015].

¹⁹⁰ <http://kops.uni-konstanz.de/> [letzter Zugriff 14.5.2015].

¹⁹¹ <http://www.ub.uni-konstanz.de/openaccess/open-access-publikationsfonds-der-universitaet-konstanz/> [letzter Zugriff 14.5.2015].

¹⁹² <http://www.ub.uni-konstanz.de/openaccess/open-access-zeitschriften/> [letzter Zugriff 14.5.2015].

¹⁹³ <https://scikon.uni-konstanz.de/home/> [letzter Zugriff 14.5.2015].

¹⁹⁴ <http://oa.uni-stuttgart.de/publizieren/fonds/index.html> [letzter Zugriff 14.5.2015].

¹⁹⁵ <http://elib.uni-stuttgart.de/opus/> [letzter Zugriff 14.5.2015].

¹⁹⁶ <http://oa.uni-stuttgart.de/oa/oa-policy.html> [letzter Zugriff 14.5.2015].

¹⁹⁷ <http://www.ub.uni-tuebingen.de/publizieren/open-access-publikationsfonds.html> [letzter Zugriff 14.5.2015].

¹⁹⁸ <https://open-journals.uni-tuebingen.de/ojs/> [letzter Zugriff 14.5.2015].

¹⁹⁹ <https://hsbiblio.uni-tuebingen.de/xmlui/> [letzter Zugriff 14.5.2015].

²⁰⁰ <https://tobias-bild.uni-tuebingen.de/> [letzter Zugriff 14.5.2015].

Mannheim Die Universität Mannheim bietet z. Z. als ein Service den Publikationsserver MADOC²⁰¹ und als Datenrepositorium MADATA²⁰² an.

Hohenheim Die Universität Hohenheim hat im März 2011 mit einer Erklärung ihre Unterstützung von Open Access angekündigt und eine Policy²⁰³ veröffentlicht. Leider wurde an der Universität das Open-Access-Publikationsfond nicht mehr verlängert und die Universität bietet zur Zeit keine (Dienst-)Leistung im Bereich Open Access an. Allerdings wird in Hohenheim der OPUS-Publikationsserver betrieben. Er ermöglicht den Mitgliedern der Universität eine kostenlose Publikation von Masterarbeiten, Dissertationen, Diskussionspapieren und Aufsätzen.

Ulm Auch an der Universität Ulm ist ein Publikationsserver (VTS²⁰⁴) und ein Open-Access-Publikationsfond²⁰⁵ eingerichtet, um Publikationen und Ausarbeitungen an der Universität frei zur Verfügung zu stellen.

In der folgenden Tabelle sind noch einmal alle entsprechenden Dienstleistungen der Universitäten abgebildet.

Viele neue Onlineveröffentlichungswege bieten die Möglichkeit, Ausarbeitungen mit interaktiven Schaubildern zu versehen, Reviewverfahren öffentlich zu gestalten und anhand eines direkten Austauschs die Kommunikation zwischen den Forschenden zu unterstützen. Auch Videos von Forschungsprozessen kann man mittlerweile in Journals publizieren. All diese Möglichkeiten scheinen sich aber wegen des übermächtigen Zitationsrankings nur langsam zu etablieren.

Standort	Pub. Server	OA-Fond	Jour. Soft.	Repo.	Info. Pla.
Freiburg	Ja	Ja	Nein	Nein	Nein
Heidelberg	Ja	Ja	Ja	Ja	Nein
Hohenheim	Ja	Nein	Nein	Ja	Ja
Karlsruhe	Ja	Ja	Ja	Nein	Nein
Konstanz	Ja	Ja	Ja	Nein	Ja
Mannheim	Ja	Nein	Nein	Ja	Nein
Stuttgart	Ja	Ja	Nein	Nein	Nein
Tübingen	Ja	Ja	Ja	Ja	Ja*
Ulm	Ja	Ja	Nein	Nein	Ja*

Tab. 13: Aktuelle Open-Science-Dienstleistungen der Universitäten (* = im Aufbau)

²⁰¹ <https://ub-madoc.bib.uni-mannheim.de/> [letzter Zugriff 14.5.2015].

²⁰² <https://madata.bib.uni-mannheim.de/> [letzter Zugriff 14.5.2015].

²⁰³ <https://kim.uni-hohenheim.de/95528> [letzter Zugriff 14.5.2015].

²⁰⁴ <http://vts.uni-ulm.de/index.asp> [letzter Zugriff 14.5.2015].

²⁰⁵ <https://www.uni-ulm.de/einrichtungen/kiz/bibliothek/wiss-informationsdienste/oap.html> [letzter Zugriff 14.5.2015].

6.5.4 E-Science-Bestrebungen

Die Ziele einer offenen Wissenschaft verfolgt auch die E-Science-Strategie des Landes Baden-Württemberg, in welche sich u. a. mit einer Open-Access-Strategie des Landes beschäftigt. Hierbei werden drei Punkte besonders hervorgehoben:²⁰⁶

- Einen größtmöglichen Nutzen aus den für die Bildung und Wissenschaft aufgewendeten Steuermitteln für die Wissenschaft, Wirtschaft und Gesellschaft zu erzielen
- Die Sichtbarkeit der wissenschaftlichen Forschung in Baden-Württemberg weltweit zu erhöhen
- Die Wettbewerbsfähigkeit des Landes nachhaltig zu steigern und zu sichern

Der „goldene Weg“ zu Open-Access-Publikationen wird auch in der E-Science-Strategie vorgeschlagen.²⁰⁷ Teilweise wird dies vorbereitend durch den sogenannten „grünen Weg“ der auf Zweitpublikationsunterstützung setzt²⁰⁸. Zusätzlich möchte das Land „Infrastrukturen mit dem Potential, eine herausragende Bedeutung für die internationale wissenschaftliche Informationsversorgung in einem besonders kompetitiven Forschungsfeld zu erlangen“, gesondert als „Open-Access-Leuchtturm“ fördern.²⁰⁹

6.5.5 Gegenwärtige Handlungsempfehlungen

6.5.5.1 Open-Source-Plattform

Der Einsatz von Open-Source-Software fördert die Kooperation der Forschenden und hilft bei der schnellen Umsetzung von einer Idee zum finalen Resultat. Zudem hilft Open Source bei der fehlerfreien Reproduktion von Ergebnissen und erspart viel Zeit. Hier könnte eine Onlineplattform bei der Vernetzung von Forschenden helfen, welche Interesse an der Lösung gleicher oder zumindest ähnlicher Probleme haben. Die Plattform könnte so gestaltet werden, dass die Forschenden eigene Programmentwicklungen vorschlagen und über andere abstimmen können. Es müsste auch in Anbetracht zu senkender Lizenzkosten überlegt werden, wie nachgenutzte wissenschaftliche Software sich für dessen Schöpfer lohnend niederschlagen könnte.

6.5.5.2 Neue Publikationsformen

Die Benutzung einer neuen und direkteren Publikationsform könnte die Zeit zur Veröffentlichung verringern. Dennoch stellt sich die Frage, ob alle neuen Möglichkeiten und auch unsere Empfehlungen in [2.3. Wissenschaftliche Datenkultur](#) das Kernproblem wirklich treffen. Immer neue Verpackungen und Anreize lösen nicht das Problem den Forschungsprozess

²⁰⁶ ESFK, 14.

²⁰⁷ ESFK, 63.

²⁰⁸ ESFK, 61.

²⁰⁹ ESFK, 63.

an sich zu veröffentlichen, wenn dafür kein System existiert.²¹⁰ Man beschränkt sich derzeit auf Forschungsprozessbeschreibungspublikationen, weil das der historisch kleinste gemeinsame Nenner ist, der theoretisch von jedem nachgenutzt werden kann. Wir denken das aktuelle Publikationswesen könnte durch ein völlig anderes Konzept zur Reputationsschätzung das in all den drei Punkten – Zeit, Kosten und Nutzen – jeweils deutlich überlegen ist, ersetzt werden. Wir haben zwar Ideen dazu, allerdings müsste so etwas von anderen im Detail ausgearbeitet und europaweit als Ersatz eingesetzt und beispielsweise durch die RDA weiter globalisiert werden.

6.6 Vorbehalte gegenüber dem Forschungsdatenmanagement (34)

6.6.1 Einleitung

Bei der Vorbereitung und Durchführung der Interviews stießen wir zuweilen auf Vorbehalte gegenüber der Zielsetzung des Projektes oder des Datenmanagements allgemein. Diese wurden auch nicht immer offen geäußert, sondern waren teilweise nur zu erahnen. Hinweise darauf sind bspw. sehr knapp und reserviert verlaufende Interviews oder ebensolche Reaktionen auf Terminanfragen. Explizite Bedenkensbekundungen und Vorbehalte in den Interviews als Antworten auf die Fragen des Interviews sind in diesem Kapitel zusammengefasst, auch wenn Vorbehalte in den Interviews vermutlich nicht immer geäußert wurden, um den Eindruck mangelnder Kooperationsbereitschaft zu vermeiden. Insgesamt liegen diesem Kapitel 34 User-Stories zugrunde.

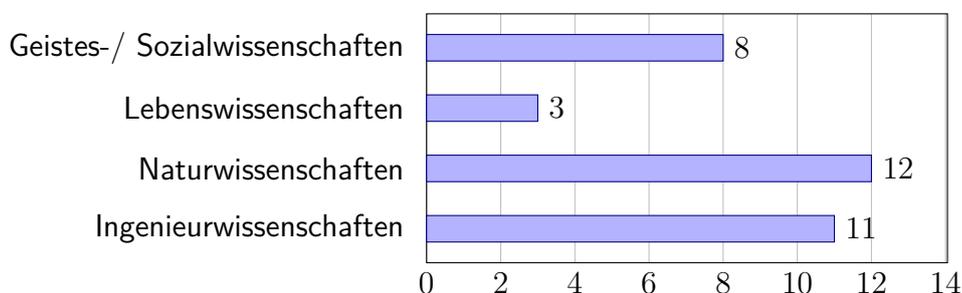


Abb. 46: Verteilung der User-Stories nach der Fachsystematik der DFG

Vier User-Stories betreffen mehrere Aspekte der angesprochenen Vorbehalte. Erkennbar wurden als wichtige Gründe für Vorbehalte folgende Befürchtungen:

- zusätzlichen Aufwand zu betreiben, durch den weniger Ressourcen für anderes verbleiben,
- Freiräume zu verlieren,

²¹⁰ In der SCM-Analogie würde der aktuelle Zustand bedeuten, dass ein Lieferant einen Rohstoff in kleinen Portionen verpackt und diese erst wieder entpackt werden müssen, was doppelten Aufwand erzeugt.

- durch Verwaltungsaufwand und Bürokratie zu sehr anderweitig belastet zu sein,
- dass Open-Access-Publikationen oder Datenpublikationscredits der Qualität schaden würden.

6.6.2 Bedarf der Wissenschaftsgemeinschaft

6.6.2.1 Aufwand und Effizienz (18)

„Als Forschende, wünschen wir uns auch beim FDM eine sorgfältige Kosten-Nutzen-Abwägung, um nicht unnötig Zeit und Geld zu vergeuden.“

Es wird beispielsweise befürchtet, dass ein allgemeines Repositorium an der Universität eingeführt wird, das keinen praktischen Nutzen hat, sondern Mehrarbeit bedeutet, weil fachspezifische Repositorien ohnehin benutzt würden. Auch soll hier den Forschenden eine Abwägung zugestanden werden, ob beispielsweise nur die Reproduzierbarkeit der Daten sichergestellt ist, wenn diese einfach zu erzeugen sind (beispielsweise durch HPC-Simulation) oder man sich auf die Existenz einer externen Kopie verlassen kann. Bei selten nachgefragten Daten kommt manchen Forschenden eine Weitergabe auf Anfrage weniger aufwendig vor als die Einstellung in ein Repositorium. Gelegentlich wurde auch über den Aufwand für Verwaltung und Bürokratie geklagt. Dies ist zwar kein spezifisches Problem des FDMs, wird aber in eine ähnliche Ecke gestellt oder als Grund angegeben, wieso für das FDM keine Zeit bliebe.

6.6.2.2 Besonderheiten und Freiräume (11)

„Als Forschende, wünschen wir uns beim FDM eine Berücksichtigung unserer fachspezifischen Eigenheiten, um negative Einflüsse auf unsere Forschung ausschließen zu können.“

Hier sehen die Forschenden bei zentral geplanten Diensten die Gefahr, nicht flexibel genug zu sein, um sich am tatsächlichen individuellen Bedarf zu orientieren. Die Erschließung eines neuen Wissenschaftsgebietes führe beispielsweise zu neuen Anforderungen bei der Erstellung von Repositorien. Die Einbindung der Forschenden in die Entwicklung von BW-Diensten sei darum erforderlich. Die Unabhängigkeit und Entscheidungsfreiheit der Wissenschaft erfordere, trotz Verbesserungen im Datenmanagement und besserem Zugang zu Daten weiterhin einen Ermessensspielraum für jeden Einzelnen, um keine „gläsernen Forscher zu schaffen, die sich ständig dokumentieren und rechtfertigen müssen“. Eine Forderung nach standardisierten Metadaten gefährde außerdem manchmal die Freiheit der Forschung, denn sie enge die Annäherung an einen Forschungsgegenstand ein²¹¹. Ein Interviewpartner wünschte sich auch in Zukunft Laborbücher in Papierform zu führen, da diese für ihn flexibler als rechnerorientierte Lösungen seien.

²¹¹ Die Aussage stammt aus einer Kulturwissenschaft. Hier seien fachspezifische und forschungsethische Aspekte beim Umgang mit Daten und die gänzlich andere Forschungslogik im Vergleich zu den Naturwissenschaften zu berücksichtigen.

6.6.2.3 Datenmenge und Qualität (5)

„Als Forschende, wünschen wir uns die (Daten-)Publikationen in unserem Fachbereich noch überschauen sowie deren Qualität vertrauen zu können, um wichtige neue Erkenntnisse unseres Fachgebiets identifizieren zu können.“

Es besteht gegenüber Open-Data-Verpflichtungen die Befürchtung, dass eine Flut von nicht publikationswürdigen Daten entsteht sowie Aufwand diese zu erzeugen und zu durchsuchen. Auch ob Open-Access-Journale (welche schließlich Geld bekommen, wenn sie einen Artikel annehmen) das richtige Mittel sind, um sorgfältige Auslese bei Publikationen zu betreiben wurde bezweifelt. Solche Ängste sind teilweise ein wenig diffus formuliert, aber auch nicht völlig von der Hand zu weisen. Auch bei derzeitigen Publikationen sei teilweise kein vernünftiger Qualitätsstandard mehr gesichert. Im Gegensatz zu den User-Storys aus [2.3. Wissenschaftliche Datenkultur](#), sehen die skeptischen Stimmen „mehr Daten“ aber nicht als Lösung, sondern befürchten dadurch eine Verschlimmerung des Problems.

6.6.3 Gegenwärtige Handlungsempfehlungen

6.6.3.1 Motivation und Information

Offensichtlich ist es wichtig, die Motivation der Forschenden zum FDM zu verbessern und unnötige Befürchtungen durch Information zu entschärfen. Bessere Informationen zu Anforderungen und Methoden des FDMs sind nicht nur notwendig, damit FDM durchgeführt werden kann, sondern sind auch erforderlich, um Ängste zu nehmen.

6.6.3.2 Bildung eines langfristigen Personalpools für FDM

Um Expertise im FDM zu erhalten, die Einzelne nach ihrem Studium erworben haben, können Anreizsysteme und permanente Stellen geschaffen werden, um sicherzustellen, dass die langfristige Perspektive der Datenhaltung und das erforderliche Inventar (Dateninfrastruktur) langfristig zur Verfügung steht.

6.6.3.3 Freiräume erhalten und Aufwand begrenzen

Weiterhin ist darauf zu achten, dass Freiräume nicht unnötig beschnitten werden, insbesondere dass verbindliche Vorgaben nicht über das notwendige Maß hinaus erfolgen oder unnötig viel Mehrarbeit erzeugen.

7 Fazit

Für eine nutzerorientierte Umsetzung von Unterstützungsprojekten bietet ein Blick in die zum jeweiligen Projekt passenden User-Stories in unseren [Daten](#)²¹² ein schnelles Verständnis für die genaue Anforderungen, um die Bedürfnisse der Forschenden bestmöglich zu berücksichtigen.

Anreize zur Datennachnutzung Betrachtet man die Gründe, wieso Forschende ihre Forschungsdaten nicht nachnutzbar zur Verfügung stellen, dann lässt es sich häufig darauf herunterbrechen, dass sie selbst davon nicht genug profitieren oder nicht wissen, wie sie profitieren können. Es ist ein systematisches Problem, dass es kaum kurzfristige Anreize für Datenbereitstellung gibt. Um Wissen von verschiedenen Forschenden wie Strom aus der Steckdose nach Belieben umleiten und verarbeiten zu können²¹³, wäre das erste Minimalziel, dass für Fachpublikationen relevante Daten nachnutzbar sind. Dafür braucht es Anreize, Klarheit über die Rechtslage der Datenweitergabe, Leitlinien, die Gewohnheit Daten bereitzustellen sowie geeignete Strukturen und womöglich neue Publikationsformen. Als Anreiz könnte die Bedeutung von Datenpublikationen bei Berufungsverfahren stärker berücksichtigt werden und Best-Practices, Informationen (auch „Werbung“ für Datenbereitstellung) und Hilfen zur Aufbereitung von Datensätzen zur Publikation angeboten werden. Auch kostenloser Speicher zur Datenveröffentlichung kleiner Datensätze mit Schnittstellen zu Standardrepositorien und geeignetere rechtliche Rahmenbedingungen wären für die allgemeine Datenverfügbarkeit förderlich. Mit anderen Maßnahmen wie (Open-Access-)Journalen für bspw. Videos oder Quellcodes könnte den neuen Anforderungen der E-Sciences begegnet werden. Auch Möglichkeiten wie offenere Reviews oder Auszeichnungen für Forschende, bspw. für „beste Datenpublikationen einer Doktorarbeit/in einer Forschungsdisziplin“, können dazu beitragen, Transparenz und Datenbereitstellung selbstverständlicher werden zu lassen.

Projekt- und Datendokumentation Oft sind Forschungsarbeiten so komplex, dass selbst beteiligte Forschende Hilfe beim Überblick über ihr Projekt brauchen. Klare Dokumentation, sowohl für einzelne Datensätze, als auch für große Forschungsprojekte, war zwar schon immer ein Kernelement wissenschaftlicher Arbeit, doch in einer immer komplexeren, schnelleren und spezialisierteren Welt, ist standardisierte Dokumentation dringlicher denn je, um die Gruppe der Menschen die ein Ergebnis nutzen können, genügend groß zu halten. Die Projektdokumentation kann ein Einstieg in die Datendokumentation sein, dient aber vor allem zur lokalen Suche und Verwaltung innerhalb einer Einrichtung. Hier könnten Standards etabliert werden, die auch in der Lehre zu vermitteln wären, um Datenbestände an Instituten geordnet ablegen zu können. Für die Leitungsebene sind Workshops und Best-Practices

²¹² Datenseite von *bwFDM-Communities*: <https://bwfdm.scc.kit.edu/daten>

²¹³ *ESFK*, 9.

wichtig, wie man Dokumentationsstrukturen in der eigenen Institution erfolgreich etabliert. Ist ein Datensatz zuordenbar, sollten Metadaten erlauben können, den Inhalt automatisiert zu erfassen und weiterzugeben. Forschende brauchen hierzu einheitliche Richtlinien und eine Infrastruktur für die Metadatenanreicherung. Das beinhaltet auch durchdachte Abläufe, sodass Daten in geeignete Repositorien münden, sowie ein breites Unterstützungsangebot, welches Expertenberatung, Software und Informationsmaterial vermitteln kann. Das Thema Metadatenanreicherung sollte schon in der Lehre verankert und möglichst anwendungsorientiert umgesetzt werden. Für letzteres sind Hilfstools zu FDM nötig, die Entscheidungshilfen (z. B. für Repositorien) bei Projektstart geben und die eine Dokumentation der Daten über die Projektdauer hinweg erleichtern. Am Projektende sollten Daten an ein geeignetes Repository übergeben werden.

Repositorien Daten sind nur nachnutzbar, wenn sie sichtbar, gut aufbereitet und unter ihresgleichen erfasst sind. So verschmelzen für Forschende Repositorien zunehmend mit Archiven. Durch einen ganzheitlichen Ansatz der Repositorienbenutzung kann, dass Potential aller wichtigen Forschungsdaten ausgenutzt werden. Die zukünftigen Repositorien sollten dabei institutionelle Zuverlässigkeit mit Interoperabilität und Benutzerfreundlichkeit verbinden. Dabei kann auf schon vorhandene Strukturen aufgebaut werden. Für Repositorien sollte konkrete Aufbauhilfe in Form von Information, Beratung und Projektförderung stattfinden, die Anforderungen wie Zugriffskontrolle, Karrenzeit, Speicherplatz, Referenzierbarkeit, Visualisierung, gute Durchsuchbarkeit oder Schnittstellen / APIs je nach Bedarf berücksichtigt. Solche Möglichkeiten bieten für Forschende neben der reinen Verfügbarmachung direkte Vorteile, was für die Verwendung von Repositorien essentiell ist. Werden Repositorien besser standardisiert, sind auch repositorienübergreifende Suchen, Verknüpfungen und eine durchgängige Repositorienkultur bei der Ablage von Forschungsdaten denkbar. Die Frage der Finanzierung von Repositorien, der Metadatenaufbereitung und Langzeitverfügbarkeit sind dabei die Grundprobleme, um neben der reinen Bereitstellung nach der Datenerzeugung auch Archivierungsaspekte abzufangen. Der hohe Return of Investment von gut geführten Langzeitrepositorien nimmt alle Akteure in die Pflicht diese Probleme zu lösen. Ein kostenloses Speicherangebot für Repositorien mit kleiner Gesamtdatenmenge, die von baden-württembergischen Einrichtungen betrieben werden und Aufbauhilfen für standardisierbare Fachrepositorien wären zwei Maßnahmen, um ein Stück voranzukommen.

Informationsangebote Weil FDM im Studium kaum berücksichtigt wird, gibt es einen hohen Informationsbedarf seitens der Forschenden, der in aller Regel auch selbstkritisch wahrgenommen wird. Allerdings übersteigt eine tiefe Expertise für FDM meistens das, was Forschende leisten können. Hier ist ein vielfacher Wunsch einfache, handhabbare und klare Leitlinien, die effizient umsetzbar sind, zu haben. Dafür wäre eine koordinierte landesweite Informationsstrategie zu entwickeln. Die Landesuniversitäten könnten gemeinsames Material zu FDM, Landesservices, Open Access, E-Science, Softwaretools, Repositorien, Fördermittelgeber usw. zusammenstellen und dabei Synergien nutzen, um praxistauglich und hochwertig, auf einer ersten gemeinsamen Ebene, Informationsbedarfe abzufangen. Dieses Material kann in Form eines Webportals, eines dazugehörigen Newsletters, Online-(Video-)Tutorials und

Broschüren angeboten werden. Auch Links zu (fach)spezifischen Expertenseiten sollten in dieses Angebot mit einfließen. Hinter einer solchen zentralen Anlaufstelle könnte ein kleines verteiltes Supportteam die zuständigen Stellen für Anliegen der Forschenden identifizieren und sie an diese vermitteln. So reduziert sich die Barriere für Forschende zur Lösung von FDM-Problemen. Den dabei gewonnenen zentralen Überblick über die landesweiten Informationsbedarfe kann das Supportteam dazu nutzen, das Informationsmaterial stetig auf die Bedürfnisse der unterschiedlichen Wissenschaftsbereiche anzupassen.

Datenaustausch und verteiltes Arbeiten Die Anforderungen an IT-Lösungen zum Datenaustausch und zu verteiltem Arbeiten mit Daten sind spürbar gewachsen. Die Möglichkeit mit entfernten Partnern große Datenpakete auszutauschen und Dokumente gemeinsam zu bearbeiten oder eigene Daten über verschiedene Systeme hinweg zu benutzen, bergen deutliches Verbesserungspotential bei Struktur, Zeiteffizienz und Produktivität. So wundert es nicht, dass personenbezogene Daten derzeit auf Systemen liegen, die dafür nicht vorgesehen sind und Forscher den Überblick über den Bearbeitungszustand (Versionierung) ihrer Daten auf den verschiedenen Systemen verlieren. Es bedarf einer groß angelegten Bereitstellung der nötigen zukunftsorientierten Infrastruktur, um zu virtuellen Forschungsumgebungen zu gelangen. Vorhandene Landeslösungen für Speicher, Computing, Datenaustausch und Authentifizierung müssen dazu miteinander kompatibel weiterentwickelt werden, sodass untereinander keine Barrieren existieren. Das beinhaltet bspw. die Verarbeitbarkeit von personenbezogenen Daten und eine Nutzerpräsentation, die wie ein einziges Dateisystem wirkt. Hier steht für die Forschenden die Benutzerfreundlichkeit im Vordergrund. Darum könnte bei Entwicklungsprojekten in Zukunft für belegbare Nutzerfreundlichkeit auch zusätzliche Förderung bereitgestellt werden. Die steigende Komplexität der Forschung macht es erforderlich, mit vielen Experten nahtlos zusammenarbeiten zu können. Dafür fehlt vor allem auch die Möglichkeiten Rechte auf Daten zu versenden (anstatt Daten selbst) und die technischen Schnittstellen und Werkzeuge, um Daten zwischen Archiv, Repositorium und Arbeitsplatz nahtlos (per Mausclick) auszutauschen und mit anderen zu teilen.

Forschungsnahe Unterstützungsangebote Auch die IT-Fähigkeiten der Forschenden bestimmen zunehmend die Umsetzbarkeit von Forschungsideen. Typische Beispiele für neue digitale Anforderungen disziplinspezifischer Forschung sind Web-Crawling, Datenkompression am Messinstrument, statistisch komplexe Analysen, das Aufsetzen einer (Web-)Datenbank oder das Entwickeln einer Modellsimulation. Die IT- bzw. Informatik-Befähigung der Forschenden sollte so hoch sein, dass sie in ihrem Gebiet relevante Möglichkeiten erkennen und mit wenig Hilfe auftretende Hürden überwinden können. Diese Hilfe sollte den Forschenden gebündelt angeboten werden, um Synergieeffekte zwischen ähnlich gelagerten Aufgaben ausnutzen zu können. Ein zentrales gut sichtbares Verzeichnis mit Spezialisten aus verschiedenen Disziplinen, die sich für gewisse Aufgabenbereiche anbieten (schwierige Visualisierung, Big-Data-Verarbeitung usw.) und im Gegenzug Anwendungsfälle für ihre Forschungsarbeit finden, könnte ohne viel Aufwand Früchte tragen. Ergänzt werden könnte so ein Informationssystem über vorhandene Expertisen durch einen kleinen Pool mit IT-Spezialisten, welche man für einzelne Aufgaben mit einem Arbeitsumfang auf Wochenskala ausleihen kann. Eine

weitere Strategie wäre der Versuch eine „Science@home“-Community aufzubauen, in welcher klar definierte IT-Aufgaben an freiwillige Laien verteilt werden. Das Projekt müsste aber zentral gesteuert und professionell überwacht werden. Für ganze Forschungsrichtungen oder Forschungsquerschnitte, die speziell gefördert werden sollen, wäre es durchaus sinnvoll, spezifisch zugeschnittene kleine IT-Labs anzubieten (siehe Data-Life-Cycle-Labs im E-Science Fachkonzept²¹⁴), in denen IT-Kräfte eng verzahnt mit Fachkräften der Forschungsdisziplin Lösungen für den ganzen Forschungszweig entwickeln.

Datenschutz und Urheberrecht Rechtliche Schwierigkeiten des FDMs liegen hauptsächlich im Urheberrecht und Datenschutz. In beiden Bereichen gibt es ein über Landesuniversitäten, ZENDAS und verschiedene Einrichtungen verteiltes Beratungs- und Informationsangebot. Hier wäre es hilfreich, unter Einbezug aller beteiligten Einrichtungen, Best-Practices und konkrete Handreichungen für die Wissenschaftler zur Verfügung zu stellen. Daneben gibt es zwar an allen Landesuniversitäten Ansprechpartner für beide Themen, dennoch scheint der Kernbedarf der FDM-Rechtsberatung oft noch ungedeckt. Äußerungen der Forschenden zu den Gründen gingen in unterschiedliche Richtungen. Eine Erklärung ist, dass Forschende nicht wissen, wo sie Hilfe finden können bzw. es ihnen nicht bewusst ist, dass sie Beratung brauchen. Eine andere Möglichkeit ist, dass die aktuelle juristische Beratung zu wenig Detailverständnis über die Forschung hat, um Interessen- und Risikoabwägung zwischen Forschungsfreiheit und Schutzinteressen vorzunehmen. In diesem Fall brauchen die Forschenden Rückendeckung und Kenntnisse die ihnen ermöglichen nicht jede Datensicherheitsempfehlung wortgetreu umsetzen zu müssen, wenn der zusätzliche Schutz gegen ein ohnehin immer existentes Restrisiko die Forschungsabläufe in außerordentlichem Maß behindert. Einfache Schrankenbestimmungen im Urheberrecht und Datenschutz würden viele Abläufe vereinfachen, in denen aktuell in jedem Einzelfall zwischen Zweck und Schutz juristisch(!) abzuwägen ist und kaum jemand existiert, der dafür wirklich Kompetenz hat. Auch gegenüber institutionellen Dritten ist die rechtliche Position des Forschenden oft unbefriedigend. Einige Forschende fühlen sich z. B. durch Journale, die durch Vertragsklauseln Ansprüche an Daten, Abbildungen und Texten erwirken wollen (oder Industriepartnern an Daten und gemeinsamen Quellcode) am kürzeren Hebel, weil sie von diesen abhängig sind. Hier könnte man anstreben, vom Bundeskartellamt stärkere Aktivität zu fordern.

Außeruniversitäre Daten Etwa 80 % der User-Stories im Bereich außeruniversitärer Daten richten sich an Behörden oder öffentliche Institutionen (z. B. Krankenkassen, Bundeswehr). Der Rest behandelt verschiedene Komplikationen bei der Erhebung aus (meist kommerziellen) Datenbanken oder Unternehmen. Die erste Kategorie betrifft direkt das Thema Open Government. Es geht also um Zugang zu Daten, die außerhalb eines Forschungskontextes erhoben wurden, aber forschungsrelevant sind (z. B. behördlich erfasste Pegelstände von Flüssen für die Wasserforschung). Die verschiedenen Wünsche der Forschenden können an die entsprechenden Stellen herangetragen werden. Allerdings gibt es, neben allgemeinem Zugang, oft auch Probleme beim Beantragungsweg, den Formaten oder der Übersichtlichkeit. Ein praxisnahes Worst-Case-Szenario sind Forschende, die zwölf Monate Kommune um Kommune

²¹⁴ ESFK, 90.

anschreiben, um zu fragen was dort für Daten zu ihrem Thema vorliegen, um dann einen Zugang zu beantragen und anzufangen alle Daten einzeln in einheitliche wissenschaftlich brauchbare Formate zu übertragen. Hier braucht es öffentliche Datenbanken der vorhandenen Datenbestände mit einheitlichen und maschinenlesbaren Formaten, standardisierten Antragsverfahren und zur Durchsetzung einen Rechtsanspruch auf die Daten in sinnvoller Form. Bei personenbezogenen Daten wird es oft noch schwieriger, diese bedarfsgerecht und reproduzierbar zu erforschen, wenn keine forschungsfragengerichtete Anonymisierung geschieht und die Primärdaten nicht zur Verfügung gestellt werden. Gegenüber der Wirtschaft wünschen sich viele Forschende mehr Vertrauen, Anreize und Bereitschaft, Forschung auch im eigenen Interesse mit Daten zu unterstützen. Externe Datenbanken, deren Zugang teuer und daher nur vereinzelt vorhanden ist, stellen die Forschung vor Probleme der gegenseitigen Nachvollziehbarkeit und langfristigen Verfügbarkeit. Sind die Daten dann auch noch in einem Spezialformat, schränkt dies die Wahrscheinlichkeit der Nachnutzung noch einmal erheblich ein. Hier müsste bei den Lizenzierungsverhandlungen gegengesteuert werden.

Formate und Standards Offene Standards bilden die Brücken zur Benutzung von Daten über verschiedene Systeme hinweg. Geräte- oder softwarespezifische Insellösungen sind ein großes Hindernis in der Wissenschaft. Es fehlt eine weit sichtbare Informations-, Beratungs- und Unterstützungszentrale für Open-Source-Software, Formate, Konvertierung und Metadaten, die auch sehr fachspezifisch Beschwerden über Inkompatibilitäten sammelt und Alternativen zu proprietären Lösungen empfiehlt. Eine bundesweite standardisierte Liste mit unterstützten Formaten von Software und Geräten könnte Druck auf Hersteller ausüben offene Formate zu unterstützen. Diese Liste könnte im Zuge einer Landesinformationsstrategie auf zentraler Stelle stehen und mit Hilfe der Forschenden gebildet werden. Das alles kann durch Richtlinien unterstützt werden, bei gleicher Eignung Lösungen mit offenen Standards zu bevorzugen. Einheitliche Softwarelizenzen im Land können die bestehenden Austauschprobleme teilweise verkleinern. Auch eine Standardisierung von Metadaten ist für maschinenles- und automatisierbare Nachnutzung unerlässlich. Für Formatfragen in Datendokumentation, Langzeitarchivierung und wissenschaftlicher Zusammenarbeit sind Repositorien ein Schlüsselement, um Standardisierung zu erreichen. Durch einen abgestimmten Repositoriumsansatz der Datenbereitstellung können allgemeine Gütekriterien für Repositorien auch Kriterien für Formate und Metadaten in den Disziplinen etablieren. Dabei sollten die Fachrepositorien untereinander einem Standard unterliegen, sodass Datenbeziehungen zwischen Repositorien automatisiert erstellbar sind.

Archivierung Archivierung allgemeiner Forschungsdaten findet derzeit auf langfristig unsicheren Füßen statt. Langzeitarchivierungsbedarfe sind derzeit oft deshalb ungedeckt, weil zusätzlich zur Bit-Erhaltung auch die langfristige Auffindbarkeit und Interpretierbarkeit gewährt sein muss. Das OAIS-Modell²¹⁵ sollte hier als Vorbild an allen Landesuniversitäten dienen, um den Forschenden Archivierungsmöglichkeiten anzubieten. Generell sind für einen

²¹⁵ Referenzmodell für ein Offenes Archiv-Informationssystem – Deutsche Übersetzung (nestor-Materialien 16, Version 2), <http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0008-2013082706> [Letzer Zugriff 25.6.2015].

erfolgreichen Archivierungsprozess Leitlinien, FDM-Planungshilfen, klare Regeln, Verantwortlichkeiten, Rechte und Workflows notwendig, sowie die technische und rechtliche Möglichkeit nicht nur Bits, sondern auch solche Software zu archivieren, die Daten wiederverwenden kann. Für die Archivierungsprozesse sollte eine Strategie gefunden werden, wie eine langfristige Zugriffs- bzw. Qualitätssicherung in einem sich wandelnden Umfeld (z. B. veraltende Datenformate) geschehen kann. Das Bemühen einer Rechtsklärung im Urheberrecht bezüglich Softwarearchivierung wäre ebenso dringlich. Machbar wäre ein allgemeines Langzeitarchiv für geringe Volumina von Forschungsdaten, das bei verbundener Datenpublikation (s. [5.2. Repositorien für die Publikation von Forschungsdaten](#)) kostenlos nutzbar ist. Dies würde das Vertrauen der Forschenden stärken, Archivierungsbemühungen an ihren Daten nicht nur zum Selbstzweck, sondern für eine lange Nutzbarkeit zu leisten. Wichtig ist aber auch, schon zu Beginn des Lebenszyklus von Daten anzusetzen und in Datenmanagementplänen die Archivierungsfrage frühzeitig zu stellen, da sonst ein enormer Mehraufwand entstehen kann, Daten überhaupt archivierungsfähig aufzubereiten.

Anhang

Abkürzungsverzeichnis

BMBF = Bundesministerium für Bildung und Forschung
DFG = Deutsche Forschungsgemeinschaft
DMP = Datenmanagementplan
DOI = Digital Object Identifier
FDM = Forschungsdatenmanagement
FIS = Forschungsinformationssystem
ESFK = E-Science – Wissenschaft unter neuen Rahmenbedingungen²¹⁶
ITB = Informationstechnikbeauftragte_r
LDSG = Landesdatenschutzgesetz
LSDF = Large Scale Data Facility
MWK = Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg
PID = Persistent Identifier
RDA = Research Data Alliance
SCM = Supply-Chain-Management
SFB = Sonderforschungsbereich
UrhG = Urheberrechtsgesetz (Gesetz über Urheberrecht und verwandte Schutzrechte)
ZENDAS = Zentrale Datenschutzstelle der baden-württembergischen Universitäten

²¹⁶ Fachkonzept zur Weiterentwicklung der wissenschaftlichen Infrastruktur in Baden-Württemberg, erarbeitet von einer durch das Wissenschaftsministerium eingesetzten Expertenkommission., https://mwk.baden-wuerttemberg.de/fileadmin/redaktion/m-mwk/intern/dateien/pdf/Forschung/066_PM_Anlage_E-Science_Web.pdf [letzter Zugriff 30.6.2015].

(1) Was möchten Sie mit Ihrer Forschung herausfinden?

(1a) Was ist das Objekt der Forschung?

(1b) Was machen Sie mit dem "Forschungsgegenstand"? (=> "Methode")

(1c) Was erzeugt diese "Methode"? (=> "Resultat", wenn Daten, was enthalten sie?)

(1d) Falls die "Schwierigkeit" der Forschung noch unklar ist: Was ist die Hauptschwierigkeit Ihrer Forschung?

(2) Aus welchen Quellen entspringen Ihre Daten?

<input type="checkbox"/> Eigene Umfragen und Interviews	<input type="checkbox"/> Eigene Abbildungsgeräte realer Objekte (Kamera, Mikroskop oder ähnliches)
<input type="checkbox"/> Andere Beobachtungen	<input type="checkbox"/> Eigenes Gerät zur Eigenschaftsmessung an Stoffen, Objekten, Umwelt, Lebewesen usw.
<input type="checkbox"/> Eigene experimentelle Versuchsaufbauten	<input type="checkbox"/> Eigene Simulationen / Softwarebenutzung
<input type="checkbox"/> Fremde (Hyper-)Textdokumente	<input type="checkbox"/> Fremde Datensätze (Tabellen, Zeitkurven, Statistiken usw.)
<input type="checkbox"/> Nicht selbst aufgenommene Bilder oder Filme	<input type="checkbox"/> Anderes

(3) Mit welchen Datenarten arbeiten Sie hauptsächlich?

<input type="checkbox"/> Bild	<input type="checkbox"/> Video
<input type="checkbox"/> Audio	<input type="checkbox"/> Text
<input type="checkbox"/> Tabelle	<input type="checkbox"/> Datenbanksystem (SQL usw.)
<input type="checkbox"/> Fachspezifisches Format (z.B. Messdaten)	<input type="checkbox"/> Fremde Datensätze (Tabellen, Zeitkurven, Statistiken usw.)
<input type="checkbox"/> Nicht selbst aufgenommene Bilder oder Filme	<input type="checkbox"/> Anderes

kurz anmerken, was der Dateiinhalt jeweils ist (z.B. Bilder von was)

Umfeld der Daten

(4) Wie viel Speicherplatz für Forschungsdaten brauchen Sie pro Jahr?

1GB 10GB 100GB 1TB 10TB 100TB

GB Wenn es eine besondere Gewichtung gibt, erwähnen: Input-, Primär-, Sekundär-, Publikationsdaten.

(5) Könnte diese Datenmenge in Zukunft wachsen? Warum? (Wachstum entsteht meist durch mehr Informationen oder höherer Genauigkeit zukünftiger Daten) Was könnten Sie dann gebrauchen?

(6) Gibt es noch etwas ganz anderes, dass Sie gerne erforschen würden, was aber zur Zeit aufgrund von nicht verfügbaren Quellen, mangelnden Ressourcen, Datenschutzgesetzen oder anderen Einschränkungen nicht machbar ist?

(7) Forschen Sie an personenbezogenen Daten? Beschränkt sich die Personenbezogenheit der Daten auf die Primär- bzw. Rohdaten? Wenn nicht, warum nicht? Wie gehen Sie aktuell damit um? Brauchen Sie Beratung/Hilfe? Wie gut funktioniert die Anonymisierung? Was ist Ihre Methode dafür?

(8) Industriepartner und (Urheber)rechtsansprüche: Welche rechtlichen Ansprüche haben Dritte an Ihren Daten?

<input type="checkbox"/> Keine	<input type="checkbox"/> Geheimhaltungsansprüche	<input type="checkbox"/> Weiß nicht
<input type="checkbox"/> Erstnutzungsrecht	<input type="checkbox"/> Andere:	

(9a) Welche Formate, die nicht offen definiert oder nur mit Spezialsoftware nutzbar sind, verwenden Sie? Warum?

(9b) Wenn Sie kommerzielle Formate nutzen: Gibt es offene Alternativen? Welche Kosten entstehen? Wie viele Lizenzen haben Sie? Hätten Sie gerne mehr? Was wünschen Sie sich hier für eine Entwicklung?

(10) Gibt es Datenformate in Ihrem Forschungszweig mit denen Sie unzufrieden sind? Wenn ja, welche und warum?

Abb. 47: Leitfaden in Form eines Fragenkataloges für die Interviews, Fragen 1 – 10

Prozess der Datenverarbeitung							
(11a) Wo gibt es bei der Verarbeitung Ihrer Forschungsdaten einen hohen technischen Ressourcenbedarf?							
(11b) Bei sehr aufwändigem Verarbeitungsschritt:	Gehen Sie vor diesem aufwändigen Arbeitsschritt noch einmal sicher, dass alle notwendigen Daten richtig sind? Wenn ja wie? <table border="1"> <tr> <td><input type="checkbox"/> Überprüfung durch Visualisierung</td> <td><input type="checkbox"/> Checksummen, Prüfskripte usw.</td> </tr> <tr> <td><input type="checkbox"/> Das geht leider nicht</td> <td><input type="checkbox"/> Anderweitig</td> </tr> </table>	<input type="checkbox"/> Überprüfung durch Visualisierung	<input type="checkbox"/> Checksummen, Prüfskripte usw.	<input type="checkbox"/> Das geht leider nicht	<input type="checkbox"/> Anderweitig		
<input type="checkbox"/> Überprüfung durch Visualisierung	<input type="checkbox"/> Checksummen, Prüfskripte usw.						
<input type="checkbox"/> Das geht leider nicht	<input type="checkbox"/> Anderweitig						
(12a) Wo müssen manuelle Verarbeitungsschritte getätigt werden?							
(12b) Falls nicht-digitale Elemente denkbar	Welche nicht-digitalen Elemente oder manuelle Digitalisierung während der Forschung gibt es?						
(13) Wenn Sie einen zusätzlichen IT-Experten zugeteilt bekämen, wofür würden Sie ihn einsetzen?							
(14) Wenn Datenmenge > 1TB oder nicht-lokale Ressourcen benutzt werden	Transferproblem ansprechen: Wo liegen die Daten und wo muss man damit arbeiten? Treffen Sie hier auf Probleme?						
(15) Gibt es während Ihres Arbeitsprozesses Daten, die Sie nicht publizieren, die aber für andere interessant sein könnten? (eventuell nach Daten aus Frage 3 fragen)							
Lagerung							
(16) Welche Programme, Dienste und Hardware nutzen Sie, um Daten zu speichern? (Backup, Cloud, ...)							
(17a) Nutzen Sie besondere Dienste oder Programme, um Ihre Daten zu verwalten?							
(17b) Wenn kein bestimmtes Verwaltungsprogramm benutzt wird	Um Forschungsdaten so mit Beschreibungen zu versehen, dass die eigene Forschungsgruppe und man selbst sich später darin zurechtfinden kann, bedarf es am ehesten: <table border="1"> <tr> <td><input type="checkbox"/> Zeit / Weniger Druck zu schnellen Resultaten</td> <td><input type="checkbox"/> Einer entsprechenden Software / Einem Standard</td> </tr> <tr> <td><input type="checkbox"/> Vorher das Wissen, welche Daten man später wirklich braucht</td> <td><input type="checkbox"/> Es gibt zu wenig Notwendigkeit, als dass sich das lohnt</td> </tr> <tr> <td><input type="checkbox"/> Sonstiges:</td> <td></td> </tr> </table>	<input type="checkbox"/> Zeit / Weniger Druck zu schnellen Resultaten	<input type="checkbox"/> Einer entsprechenden Software / Einem Standard	<input type="checkbox"/> Vorher das Wissen, welche Daten man später wirklich braucht	<input type="checkbox"/> Es gibt zu wenig Notwendigkeit, als dass sich das lohnt	<input type="checkbox"/> Sonstiges:	
<input type="checkbox"/> Zeit / Weniger Druck zu schnellen Resultaten	<input type="checkbox"/> Einer entsprechenden Software / Einem Standard						
<input type="checkbox"/> Vorher das Wissen, welche Daten man später wirklich braucht	<input type="checkbox"/> Es gibt zu wenig Notwendigkeit, als dass sich das lohnt						
<input type="checkbox"/> Sonstiges:							
(18) Gibt es ein Standardverfahren, was nach Projektende mit Ihren verschiedenen Daten geschehen soll? Wenn nicht, planen Sie etwas?							
<input type="checkbox"/> Kein bestimmter Plan	<input type="checkbox"/> Für Rohdaten haben wir einen Ort:	<input type="checkbox"/> Darum kümmert sich ein Projektpartner					
<input type="checkbox"/> Unser Standardverfahren ist:	<input type="checkbox"/> Unser Plan ist zukünftig, dass:	<input type="checkbox"/> Alles kommt an folgenden Ort:					
(19a) Würden Sie eine Langzeitarchivierung (>10 Jahre) eines Teils Ihrer Daten für sinnvoll halten? Wenn ja, wofür wäre es Ihnen besonders wichtig?							
<input type="checkbox"/> Beweisbarkeit der Resultate	<input type="checkbox"/> Zum Vergleich mit ähnlichen Daten						
<input type="checkbox"/> Informationsquelle für zukünftige Forschung	<input type="checkbox"/> Späterer Eigenbedarf						
<input type="checkbox"/> Andere Zwecke:							
(19b) Falls "Messungen" (auch Umfragen) an "realen Dingen" oder einmalige Beobachtungen vorgenommen werden :	Könnte das für jemanden in 100 Jahren interessant sein? Speichern Sie diese Daten noch einmal extra gesichert ab?						

Abb. 48: Leitfaden in Form eines Fragenkataloges für die Interviews, Fragen 11 – 19

Publikation und "Data Sharing"		
(20a) In welchem Rahmen haben Sie schon eigene Primärdaten ausgetauscht?		
<input type="checkbox"/> In der Forschungsgruppe	<input type="checkbox"/> Innerhalb eines Projektes	<input type="checkbox"/> Auf Anfrage von Außerhalb
<input type="checkbox"/> Öffentliche Bereitstellung zum Download	<input type="checkbox"/> kaum / gar nicht	<input type="checkbox"/> Anders :
(20b) Wenn Austausch:	In welchem Zusammenhang? Welche Programme oder Dienste nutzen Sie dafür?	
(20c) Wenn kaum Austausch:	Gibt es Partner oder Kollegen mit denen Sie den Informationsaustausch gern verbessern würden?	
(21) Sind Sie mit den offen zugänglichen Forschungsdaten anderer Gruppen Ihres Forschungsbereichs zufrieden? Wo sehen Sie Verbesserungsbedarf?		
<input type="checkbox"/> Ja <input type="checkbox"/> Nein		
(22a) Welches Medium nutzen Sie zur Veröffentlichung Ihrer Ergebnisdaten?		
<input type="checkbox"/> Anhang, "weiterführendes Material" oder Hostingmöglichkeit beim Fachblatt	<input type="checkbox"/> Eigene Webseiten für das Anschauen / Herunterladen	
<input type="checkbox"/> Fachspezifisches (offenes) Datenarchiv	<input type="checkbox"/> Allgemeine Datenpublikationsmöglichkeit (z.B. an Universität)	
<input type="checkbox"/> Kein Bedarf	<input type="checkbox"/> Keine Möglichkeit	
<input type="checkbox"/> Anderes		
(22b) Falls kein Repository erwähnt wurde :	Warum benutzen Sie kein offen zugängliches Archiv für Ihre Ergebnisdaten?	
(22c) Wenn "Aufwand" ein Grund gegen Repository war:	Welche Archivierungshilfen oder andere Dienste brauchen Sie?	
(22d) Falls "Zugriffsvorbehalte" gegen Repositorien bestehen	Der Zugriff auf archivierte Daten ist unterschiedlich regelbar. Welcher Gruppe möchten Sie Ihre Daten zur Verfügung stellen (auch international)?	
Publikation und "Data Sharing"		
(23) "Sackgassen" oder "negative Befunde" werden selten publiziert, weshalb andere Forscher sie oft mühsam reproduzieren müssen. Wie sehr könnten Sie in Ihrem Gebiet von einem Überblick zu solchen Erfahrungen profitieren?		
<input type="checkbox"/> nicht so sehr <input type="checkbox"/> einigermaßen <input type="checkbox"/> gut <input type="checkbox"/> sehr deutlich <input type="checkbox"/> gar nicht einschätzbar		
(24) Was würden Sie noch brauchen, um Ihre Daten effizient und einfach veröffentlichen zu können?		
Allgemeines		
(25) Wie hoch ist der Bedarf Ihrer Forschungsgruppe, mehr über Methoden des Forschungsdatenmanagements zu erfahren? Welche Methode der Information würden Sie bevorzugen?		
<input type="checkbox"/> Dazu haben wir keine Zeit	<input type="checkbox"/> Wenn ich gutes Material bekäme, würde ich es verteilen	
<input type="checkbox"/> Ich würde mir gerne ihren Newsletter anschauen	<input type="checkbox"/> Webseite / Wiki für Forschungsdaten	
<input type="checkbox"/> Schulungen Workshops	<input type="checkbox"/> Weiß nicht	
<input type="checkbox"/> Anderes:		
(26) Was denken Sie, wie gut Sie bisher über (Landes)datendienste informiert werden? Wie möchten Sie in Zukunft über existierende oder geplante Services informiert werden?		
(27) Wenn Sie drei Wünsche zum Thema Forschungsdaten frei hätten, welche wären das?		

Abb. 49: Leitfaden in Form eines Fragenkataloges für die Interviews, Fragen 20–27