

Karlheinz Pappenberger*

bwFDM-Communities – Wissenschaftliches Datenmanagement an den Universitäten Baden-Württembergs

DOI 10.1515/bfp-2016-0017

Zusammenfassung: Im Rahmen der E-Science-Strategie des Landes Baden-Württemberg fand in den Jahren 2014/2015 eine umfassende Befragung der Wissenschaftler an den neun Universitäten des Landes zu ihrem Umgang mit Daten und zum Forschungsdatenmanagement statt. Die Ergebnisse zeigen die Relevanz des Themas auf und die vielschichtigen Bedarfe an einer konzeptionellen, inhaltlichen und technischen Unterstützung der Forscher durch Infrastruktureinrichtungen im Umgang mit Forschungsdaten.

Schlüsselwörter: Forschungsdaten; Forschungsdatenmanagement; Befragung

bwFDM-Communities – Research Data Management by Researchers in Baden-Wuerttemberg

Abstract: As part of the e-science strategy of Baden-Wuerttemberg a comprehensive survey on the topic of research data management amongst researchers from the state's nine universities was conducted in 2014/15. The results indicate that the issue is highly relevant as research communities need conceptional, technological, and consultative support from infrastructural facilities in their handling of data.

Keywords: Research Data; Research Data Management; Survey

1 Einleitung

„Eine gut funktionierende und innovative Informationsinfrastruktur ist Voraussetzung für den Erfolg der wissenschaftlichen Einrichtungen des Landes im nationalen und internationalen Wettbewerb, für Spitzenforschung und Ex-

zellenz.“¹ So benennt das Ministerium für Wissenschaft, Forschung und Kunst des Landes Baden-Württemberg auf seiner Webseite zu E-Science die Bedeutung, die es damit auch den Infrastruktureinrichtungen der Universitäten des Landes Baden-Württemberg im Forschungswettbewerb zumisst. Konkretisiert wurden die zu schaffenden Informationsinfrastrukturen im „Fachkonzept E-Science“,² einem 120-seitigen Papier des Ministeriums, das fünf zentrale Handlungsfelder benennt und dort Handlungsoptionen aufzeigt.

Diese fünf Handlungsfelder beinhalten die Lizenzierung elektronischer Informationsmedien, die Digitalisierung, Open Access, Virtuelle Forschungsumgebungen und eben das Forschungsdatenmanagement. Im Rahmen des Forschungsdatenmanagements sind grundlegende Handlungsbedarfe offensichtlich: Es geht um die nachhaltige Sicherung, Erschließung und Bereitstellung von wissenschaftlichen Daten mit dem Ziel, die Nachvollziehbarkeit dieser Daten und der daraus abgeleiteten Forschungsergebnisse zu verbessern und eine breite Nachnutzung von häufig nur einmalig generierbaren Daten zu ermöglichen. Um dieses Ziel zu erreichen, sollen im Land Baden-Württemberg entsprechende Infrastrukturen ausgebaut und forschungsunterstützende Services eingerichtet werden. Eine zehnköpfige Arbeitsgruppe aus Vertretern von Bibliotheken, Rechenzentren, Datenzentren und Wissenschaftlern unter Vorsitz des Ministeriums arbeitete dabei für den Bereich Forschungsdatenmanagement in dem Fachkonzept eine Reihe von Empfehlungen aus.³

Gerade im Bereich des Forschungsdatenmanagements sind Anforderungen an den Umgang mit Daten und die mit

1 <https://mwk.baden-wuerttemberg.de/de/forschung/forschungslandschaft/e-science/> (letzter Zugriff: 13.11.2015).

2 Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg (Hrsg.): E-Science. Wissenschaft unter neuen Rahmenbedingungen. Fachkonzept zur Weiterentwicklung der wissenschaftlichen Infrastruktur in Baden-Württemberg. 2014: https://mwk.baden-wuerttemberg.de/fileadmin/redaktion/m-mwk/intern/dateien/pdf/Forschung/066_PM_Anlage_E-Science_Web.pdf, 29.7.2014 (letzter Zugriff: 13.11.2015).

3 Ministerium für Wissenschaft, Forschung und Kunst (Anm. 2) S. 78–95.

*Kontaktperson: Karlheinz Pappenberger, Karlheinz.Pappenberger@uni-konstanz.de

einem Forschungsdatenmanagement zu erreichenden Ziele sehr heterogen, sowohl fachspezifisch als auch innerhalb einer Fachcommunity. Allgemeine Dienste und Dienstleistungen sind daher zwar unerlässlich, decken aber nur einen Teil der von den Wissenschaftlern gestellten Anforderungen ab. Die Leiter der universitären Rechenzentren in Baden-Württemberg (ALWR) haben aufgrund dieser Einschätzung Anfang 2014 das Forschungsdatenprojekt bwFDM-Communities im Land Baden-Württemberg initiiert, das sich von den seit einigen Jahren von den Rechenzentren im Land entwickelten kooperativen Projekten bzw. Diensten unterscheidet. Während die meisten Projekte eher hard- und softwareorientiert sind, ist dieses Projekt ein „Erkundungsprojekt“.⁴

2 Das Projekt bwFDM-Communities

Der Kern des Projekts bwFDM-Communities⁵ ist eine Befragung von Wissenschaftlern an allen neun baden-württembergischen Universitäten zum Thema Forschungsdatenmanagement. National und international gab es in den vergangenen Jahren eine Reihe von Umfragen, die auf lokaler institutioneller Ebene durchgeführt wurden.⁶ Im Projekt bwFDM-Communities wurde die Befragung jedoch deutlich größer und umfassender angelegt. Im Projektzeit-

raum von Januar 2014 bis Juni 2015 wurden 627 persönliche Interviews mit 779 einzelnen Wissenschaftlern zu ihrem Umgang mit Forschungsdaten durchgeführt. Es wurde hierfür ein Leitfaden in Form eines Fragenkataloges mit 40 offenen, halboffenen und geschlossenen Fragen⁷ entwickelt und dieser in durchschnittlich einstündigen Interviews mit Antworten gefüllt. Aus diesen Antworten wurden sogenannte User-Stories extrahiert, die in der späteren Analyse zu Meta-User-Stories und Themenblöcken zusammengefasst wurden. Aus der Grundgesamtheit von etwa 30 000 Wissenschaftlern an den neun baden-württembergischen Universitäten wurde versucht, ein möglichst breites Bild des Umgangs mit Forschungsdaten zu erheben, indem im Projekt möglichst alle Arbeitsgruppen und Professuren kontaktiert und um Mitwirkung bei der Befragung gebeten wurden. Die tatsächlich zustande gekommenen Interviews ergaben sich aus den faktischen Gesprächsmöglichkeiten an den einzelnen Standorten und führten damit im statistischen Sinne zu einer nicht-repräsentativen Erhebung. Einzelne Interviews wurden teilstandardisiert durchgeführt, aufgrund der Heterogenität von Forschungsdaten gab es bewusst Freiräume für die Interviewten, eigene Punkte hervorzuheben und Themenschwerpunkte zu setzen. An jeder der Universitäten Freiburg, Heidelberg, Hohenheim, Konstanz, Mannheim, Stuttgart, Tübingen, Ulm und des KIT (Karlsruhe) war für die Durchführung und Auswertung der Interviews ein Key-Accounter verantwortlich.⁸ Die Projektleitung und Projektkoordination waren am Steinbuch Centre für Computing (SCC) des KIT angesiedelt.⁹

Aus den geführten Interviews wurden insgesamt 2554 sog. User-Stories¹⁰ herausgearbeitet, wobei mit 795 User-Stories aus den Geistes- und Sozialwissenschaften, 703 aus den Lebenswissenschaften, 467 aus den Naturwissenschaften und 589 aus den Ingenieurwissenschaften eine

⁴ Kooperationen unter den Rechenzentren des Landes Baden-Württemberg werden auf einer gemeinsamen Internetseite vorgestellt: <http://www.alwr-bw.de/kooperationen/> (letzter Zugriff: 10.11.2015).

⁵ <http://bwfdm.scc.kit.edu/>. Auf der Projektwebseite gibt es auch einen Bereich „Offenes Ergebnismaterial“, das den Interviewleitfaden und einzelne Ergebnisse in Schaubildern für alle einsehbar zur Verfügung stellt. Die User-Stories sind unter <https://bwfdm.scc.kit.edu/cgi-bin/daten> in anonymer Form offen abgelegt.

⁶ Etwa die bekannte Umfrage an der HU Berlin: Simukovic, Elena; Kindling, Maxi; Schirnbacher, Peter (2013): Umfrage zum Umgang mit digitalen Forschungsdaten an der Humboldt-Universität zu Berlin. Umfragebericht, Version 1.0. Online verfügbar auf dem edoc-Server der Humboldt-Universität zu Berlin. urn:nbn:de:kobv:11-100213001. Es gibt darüber hinaus eine Vielzahl weiterer Umfragen an Universitäten zum Forschungsdatenmanagement. So etwa an der Universität Hamburg 2014: <https://www.gwiss.uni-hamburg.de/gwin/service/auswertung.pdf> (letzter Zugriff: 13.11.2015) oder an der Universität Kiel: <http://www.uni-kiel.de/vfu/de/forschungsdatenmanagement/umfrage-zum-umgang-mit-forschungsdaten> (letzter Zugriff: 13.11.2015). International lässt sich z. B. auf Umfragen aus dem Jahr 2012 an den Universitäten Oxford und Nottingham verweisen: <https://blogs.it.ox.ac.uk/damaro/2013/01/03/university-of-oxford-research-data-management-survey-2012-the-results/> (letzter Zugriff: 13.11.2015) bzw. Parsons, Thomas; Grimshaw, Shirley; Williamson, Laurian (2013): Research Data Management Survey: Report. Project Report. The University of Nottingham, Nottingham: <http://eprints.nottingham.ac.uk/1893/>.

⁷ Gefragt wurde z. B. „Wie viel Speicherplatz für Forschungsdaten brauchen Sie pro Jahr“ oder „Welche Programme, Dienste und Hardware nutzen Sie, um Daten zu speichern? (Backup, Cloud, ...)“. Der vollständige Fragenkatalog ist unter https://bwfdm.scc.kit.edu/downloads/bwFDM-Communities_Interviewleitfaden.pdf hinterlegt.

⁸ Freiburg: Dennis Wehrle; Heidelberg: Jonas Kratzke; Hohenheim: Dr. Denis Shcherbakov; Konstanz: Jessica Rex; Mannheim: Johannes Knopp; Stuttgart: Dr. Jörg Hertzler; Tübingen: Fabian Schwabe, Dr. Dieta-Frauke Svoboda; KIT: Uğur Çayoğlu.

⁹ Frank Tristram als Projektkoordinator und Prof. Dr. Achim Streit als Projektverantwortlichem.

¹⁰ Als User-Story werden die strukturierten Bedarfsäußerungen der befragten Personen verstanden, wie etwa die folgende: „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.“

breite Verteilung über die verschiedenen Fachdisziplinen sichtbar wird. Dieser User-Stories stellen die Essenz der Befragung dar und liefern breite Erkenntnisse über den Umgang mit Forschungsdaten an baden-württembergischen Universitäten, aus denen Ansatzpunkte für geeignete infrastrukturelle und politische Maßnahmen zu einem verbesserten Forschungsdatenmanagement abgeleitet werden können.

3 Vielfältige Infrastrukturbedarfe – die Projektergebnisse

Bei der Auswertung der Befragungsergebnisse überrascht die Vielfalt an Themen und Bedarfen zum Forschungsdatenmanagement, die angesprochen wurden. Die Ergebnisse gehen dabei grundsätzlich einher mit den Resultaten der benannten Befragungen zum Forschungsdatenmanagement, verdeutlichen aber aufgrund der Vielzahl der detaillierten Interviews nochmals verstärkt die Heterogenität im Umgang mit Forschungsdaten, sowohl zwischen einzelnen Fachcommunities als auch innerhalb einer Fachcommunity. Dennoch sind natürlich nicht alle 2554 User-Stories unterschiedlich. Es lassen sich durch die vielen dokumentierten User-Stories sehr gut Cluster von Bedarfen herausarbeiten, die von vielerlei Seiten immer wieder angesprochen wurden und die damit besonders zügig angegangen werden sollten.

3.1 Strukturelle Rahmenbedingungen verbessern

Eines der auffälligsten Ergebnisse der Befragung ist, wie häufig Bedarfe und Themen benannt werden, die sich unter der Überschrift „Strukturelle Rahmenbedingungen“ zusammenfassen lassen und die in direktem Zusammenhang mit der Heterogenität der Forschungsdatenlandschaft stehen. Dies war insofern überraschend, als der zugrunde liegende Fragenkatalog zwar bewusst sehr offen gestaltet war, aber doch viele Fragen zu technischen Bedarfen und der Weiterverarbeitung und Archivierung von Daten enthielt. Das Thema Informationsangebote zum Forschungsdatenmanagement nimmt dabei die gewichtigste Rolle ein und verdeutlicht die Dringlichkeit einer effizienteren Informationspolitik. In allen Fachcommunities wird gleichermaßen häufig der Wunsch geäußert, mehr über Methoden des Forschungsdatenmanagements zu wissen. Zentrale Infrastrukturdienste sind zu wenig bekannt und werden zu wenig beworben. Von großer Bedeutung in den

User-Stories sind daher auch Informationsdienste,¹¹ die Abhilfe schaffen und Informationsdefizite beseitigen können: Beratung und ein direkter Ansprechpartner für Fragen des Forschungsdatenmanagements, eine Informationsplattform, Kurse und Tutorials und Newsletter/E-Mails.

Urheberrecht und Datenschutz sind juristische Bereiche, mit den sich Wissenschaftler möglichst nicht – zumindest nicht zeitintensiv – auseinandersetzen wollen. Eine Reihe von User-Stories greift diese rechtlichen Fragen auf – mit einem Schwerpunkt bei den Geistes- und Sozialwissenschaften – und Forscher wünschen sich vor allem Unterstützung bei der Klärung von Rechtsfragen hinsichtlich der eigenen Daten und der Verwendung der Daten Dritter.¹²

Obwohl nicht direkt in der Befragung angesprochen, war eine mangelhafte wissenschaftliche Datenkultur – verteilt über alle Fachdisziplinen – ebenfalls Thema. Von vielen Forschern wurde der Wunsch nach einem Anreizsystem für Datenpublikationen und einem tiefgehenden Wandel in der Handhabung von Daten und deren Wertschätzung, auch was die Publikation von Negativbefunden betrifft, geäußert. Dass neben einem Wertewandel weitere Faktoren für ein erfolgreiches Datenmanagement notwendig sind und bereitgestellt werden müssten, wurde ebenfalls häufig angesprochen: Dazu zählen z. B. ein vorhandenes Repositorium zur Datenablage, Zeit, Geld, Personal, Beratung und Rechtssicherheit.

3.2 Effektivere digitale Forschungsdatenverarbeitung

Ein sehr offensichtlicher und nach den strukturellen Rahmenbedingungen zweiter großer Bereich auf der Wunschliste der Forscher ist die konzeptionelle und praktische Unterstützung bei der digitalen Verarbeitung von Forschungsdaten. Viele Wissenschaftler drücken den Wunsch nach einer fachspezifischen bzw. fachübergreifenden IT-Unterstützung aus, die beispielsweise von der Datenkompression und Formatkonvertierung über eine automatisierte Bildanalyse bis zur Unterstützung bei statistischen Analysen reicht. Man stellt hier eine leichte Häufung in den

¹¹ Wie z. B.: „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.“

¹² Eine typische User-Story lautet: „Als Forschende wünschen wir uns rechtliche Hilfsmittel und Beratungsangebote, um unsere Fragen zu den Themen Urheberrecht und Datenschutz zu klären.“

Geistes- und Sozialwissenschaften fest und speziell in den Digital Humanities, wo zunehmend Aufgaben mit IT-Kenntnissen anfallen. IT-Kenntnisse sind auch nötig, um wissenschaftliche Softwaretools beispielsweise zur Textanalyse, Modellierung, Simulation oder Visualisierung für Forschungsvorhaben richtig auszuwählen, zu beschaffen und effektiv einzusetzen. Ebenfalls Thema war der Wunsch nach Unterstützung beim Einsatz von Software, vor allem bei den Ingenieur- und Lebenswissenschaften. Einheitliche Formate und Standards erleichtern die effektive digitale Verarbeitung von Forschungsdaten und den Datenaustausch immens. Fehlende einheitliche Formate und Standards wurden als Hemmnis für effektives Forschungsdatenmanagement hervorgehoben, insbesondere von Natur- und Ingenieurwissenschaftlern. Diese zu beseitigen und Richtlinien zu entwickeln, wird als eine wichtige Aufgabe für Infrastruktureinrichtungen gesehen.¹³

3.3 Datenerhebung, -beschaffung und -austausch erleichtern

Eine Reihe von Forschern, vor allem aus den Geistes- und Sozialwissenschaften, wünscht sich eine Verbesserung der Nutzung von Daten Dritter¹⁴ und einen leichteren Weg, um an Daten außeruniversitärer Einrichtungen zu gelangen bzw. vorhandene Daten leichter finden zu können. Insbesondere aus dem Fachbereich Geschichte wird der Wunsch nach verstärkter Digitalisierung vor allem von Quellenmaterial angesprochen. Hinsichtlich des Austausches von Informationen und Daten mit externen nationalen und internationalen Partnern wird der Wunsch nach einfachen und praktikablen Austauschdiensten geäußert, welche den oft unpraktischen oder unsicheren Datenaustausch via E-Mail, der Ablage auf Servern, Dropbox oder USB-Stick ergänzen oder ersetzen sollen.

¹³ Einheitliche Standards waren ein häufig genannter Wunsch: „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.“

¹⁴ Insbesondere bei Daten öffentlicher Einrichtungen wird erwartet, dass man einen besseren Zugang erhält, wie es in dieser User-Story formuliert ist: „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.“

3.4 Langfristige Bewahrung von Informationen sicherstellen

Häufig wurde ebenfalls der Wunsch nach Repositorien für die Publikation von Forschungsdaten benannt, wobei vor allem fachspezifische Datenrepositorien im Vordergrund stehen neben fachlich übergreifenden Datenrepositorien und einem von einer zentralen Infrastruktureinrichtung betriebenen allgemeinen Datenrepository. Das Stichwort Archivierung¹⁵ fiel häufig in den Lebens- und Ingenieurwissenschaften, in denen es vor allem um das Problem großer Datenmengen geht, während in den Sozialwissenschaften die strukturierte Ablage von großen Umfrageergebnissen in Datenzentren eine wichtige Rolle spielt. Vor allem bei kleinteiligen und vielfältigen Daten spielt eine gute Projekt- und Datendokumentation einschließlich einer Anreicherung mit Metadaten eine wichtige Rolle, die mangels personeller Kapazitäten oft nicht umgesetzt werden kann.

3.5 Technische und personelle Ressourcen verbessern

Vor allem in den datenintensiven Lebens- und Ingenieurwissenschaften wird der Mangel an personellen Kapazitäten und technischer Infrastruktur für ein effektives Datenmanagement hervorgehoben. Die Wissenschaftler thematisieren u. a. den Bedarf an einer verbesserten Datenspeicherung, mehr Rechenleistung und einer verbesserten Geräteausstattung. Zudem wird der Wunsch nach mehr IT-Support benannt, um sich als Wissenschaftler stärker auf die eigene Forschung konzentrieren zu können.

4 Fazit

Die Befragung stieß bei den angesprochenen Wissenschaftlern auf große und breite Resonanz – einzelne Interviews dauerten bis zu vier Stunden und es hätten weit mehr Interviews geführt werden können als aufgrund der gegebenen Personalkapazitäten tatsächlich geführt wurden. Dies zeigt die hohe Relevanz und Dringlichkeit des Themas Forschungsdaten auf. Forscher fühlen sich häufig bei der Lösungsfindung bei den organisatorischen, tech-

¹⁵ User-Stories drehen sich hier immer wieder um Zugriffsregelungen: „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.“

nischen und analytischen Herausforderungen beim Umgang mit ihren Forschungsdaten allein gelassen. Sie haben neben dem konkreten Bedarf an technischen Werkzeugen vor allem auch Bedarf an verbesserten Rahmenbedingungen und kompetenter Beratung in allen Phasen des Managements von Forschungsdaten. Die Befragungsergebnisse ermöglichen im Vergleich mit dem Status Quo das Aufzeigen von Handlungsempfehlungen zur Unterstützung der Wissenschaftler für ein besseres Forschungsdatenmanagement. Der für das Ministerium für Wissenschaft, Forschung und Kunst in Baden-Württemberg verfasste Projektbericht¹⁶ enthält solche Handlungsempfehlungen, die bei den aktuellen Förderprogrammen dazu

beitragen, zielgenauer Infrastrukturprojekte auf Landesebene zur Beseitigung erkannter Defizite im Forschungsdatenmanagement zu initiieren. Ebenso wichtig sind bei global vernetzter Forschung aber auch nationale und internationale Anstrengungen und Initiativen zum Forschungsdatenmanagement und ein effektives Zusammenspiel der auf lokaler, regionaler, nationaler und internationaler Ebene entwickelten Dienste und Services.



Karlheinz Pappenberger

Universität Konstanz
Kommunikations-, Informations-, Medien-
zentrum (KIM)
Bibliotheksdienste
Universitätsstr. 10
D-78464 Konstanz
Karlheinz.Pappenberger@uni-konstanz.de

¹⁶ <http://bwfdm.scc.kit.edu/downloads/Abschlussbericht.pdf> (letzter Zugriff: 28.02.2016).