

Am Thema Forschungsdaten kommt niemand mehr vorbei

Von Jasmin Schmitz, Düsseldorf, schmitz-jasmin@web.de

Der 104. Bibliothekartag fand in diesem Jahr in Nürnberg statt. In Sessions, die sich mit der Arbeit von wissenschaftlichen Bibliotheken beschäftigten, standen in erster Linie das Management von Forschungsdaten, Infrastrukturen und Open Access im Vordergrund.

Forschungsdaten und Forschungsdatenmanagement. Sehr viele Beiträge beschäftigten sich mit dem Thema Forschungsdaten. Sie machten deutlich, dass Hochschulen und Forschungseinrichtungen eine Verantwortung für ihre Wissenschaftler haben, dass diese nämlich Forschungsdatenmanagement gemäß den Vorgaben von Mittelgebern betreiben und die Vorteile der Nachnutzung von Daten nutzen können. Hier sind vor allem Präsidium/Rektorat und sonstige Leitungsebenen gefragt, um mittels ihrer Policies einen fördernden Handlungsrahmen zu schaffen. Die finanziellen, technischen, rechtlichen und administrativen Anforderungen an diese Aufgabe sind sehr hoch. Es ist eine Zusammenarbeit zwischen Administration, Bibliotheken und Rechenzentrum erforderlich. Eine einrichtungsweite Gesamtstrategie ist hilfreich, weil dies bei Drittmittelanträgen überzeugender wirkt¹. Eine enge Zusammenarbeit mit den Wissenschaftlern der jeweiligen Disziplinen ist gleichfalls unabdingbar, weil jede Disziplin ihre eigene Definition von Forschungsdaten hat und damit auch andere Handlungserfordernisse². Forschungsdaten gerade an Hochschulen sind heterogen. Zusätzlich zu den am häufigsten vorkommenden experimentellen Daten und quantitativen Messdaten wurden in einer Umfrage der Universität Münster³ Statistiken und qualitative Befragungsdaten genannt. Nicht nur Text- und Bildformate, auch Audio- und Videodateien, Statistikformate und gerätespezifische Dateiformate sind üblich. Forschungsdaten werden noch immer auf lokalen und dienstlichen Rech-

nern und externen Datenträgern gespeichert; außerdem werden Cloud-Dienste genutzt. Nach den Regeln der guten wissenschaftlichen Praxis müssen Forschungsdaten zehn Jahre und länger gespeichert werden. Diese Vorgabe ist vielen nicht bekannt, so dass häufig kürzere Aufbewahrungsfristen gewählt werden. Lediglich in den Natur- und Lebenswissenschaften lässt sich ein gewisses Bewusstsein für angemessene Aufbewahrungsfristen erkennen. Die Hauptgründe für eine Aufbewahrung sind der Nachweis der Replizierbarkeit und Reanalysen, etwa 27% nutzen darüber hinaus die Daten zu Übungszwecken für die Lehre. Die hauptsächlichen Gründe, warum Forschungsdaten nicht zugänglich gemacht werden, lauten: rechtliche Einschränkungen (50%), fehlende Eignung der Daten (48%), fehlende Plattform (24%), Zeitmangel (17%). In den Lebenswissenschaften geben 27% der Wissenschaftler an, dass sie über geringe bis sehr geringe Kenntnisse im Datenmanagement verfügen, in den übrigen Fächern liegt der Anteil sogar zwischen 42 und 48%. Viele Wissenschaftler wünschen sich Beratungsleistungen zu rechtlichen und technischen Fragen, zur Beantragung von Drittmitteln sowie zum Publizieren und Zitieren. Bibliotheken können hier ihre Kompetenzen sowie ihr Wissen über die Sicherstellung der Wiederauffindbarkeit einbringen. Die Ausbildung von Data Librarians wird befürwortet. Bislang ist aber eher unklar, wie und wo eine solche Ausbildung stattfinden könnte⁴.

Auffindbarkeit von digitalen Objekten durch persistente Identifikatoren (PI). Für die dauerhafte Auffindbarkeit und Zitierfähigkeit digitaler Objekte sind Persistente Identifikatoren von zentraler Bedeutung. Für Volltexte und Forschungsdaten hat sich der Digital Object Identifier (DOI) durchgesetzt. Zunehmend gewinnt auch die Vergabe von Identifikatoren für einzelne Abschnitte oder Teile eines Werkes an Bedeutung. Für Filme gibt es hier-

zu den Media Fragment Identifier (MFID), der ein sekundengenaueres Zitieren ermöglicht. Die Wissenschaftsgemeinschaft ist sich einig, dass jedes Dokument oder Objekt, welches zitiert werden könnte, eine DOI bekommen sollte. Inwieweit ein Dokument zitierwürdig ist, muss die jeweilige Fachgemeinschaft entscheiden bzw. das zuständige Datenzentrum, welches Identifikatoren vergibt. Bei einer gemeinsamen Publikation von Volltext und Forschungsdaten hat sich in der Praxis durchgesetzt, dass beide jeweils eine eigene DOI erhalten, also getrennt referenzierbar sind⁵. Komplexer wird das Thema, wenn PI für dynamische Daten, also nicht-abgeschlossene Datensätze, Aggregationen und Selektionen vergeben werden sollen. Ein Beispiel für Aggregationen sind Zeitschriften, die nicht abgeschlossen sind, weil laufend neue Artikel hinzukommen. Bei Selektionen handelt es sich um temporäre Aggregationen wie z.B. Queries, deren Ergebnisse nicht stabil sind. Eine Möglichkeit für den Umgang mit dynamischen Daten stellt die Versionierung dar. Alte Versionen werden in der Regel beibehalten. Problematisch wird das Vorgehen nur, wenn es viele Versionen gibt. Zudem sind diese nicht für alle dynamischen Daten (z.B. Netzwerke) geeignet. Eine Alternative ist der Zeitstempel, der den Zeitpunkt der Änderung markiert. Dieser ist flexibler anwendbar, auch für Aggregationen⁶.

Persistente Identifikatoren für Kulturgüter. Außerhalb der Bibliothekswelt ist das Verständnis für PI noch nicht weit verbreitet. Es ist jedoch sinnvoll, Kulturobjekte aller Art mit einem persistenten Identifikator zu versehen, auch solche, die nicht digitalisiert sind, um die Auffindbarkeit beispielsweise von Objekten in Museen zu verbessern. Im Falle nicht digitalisierter Objekten würden dann zumindest Metadaten zur Verfügung stehen. Hierzu wurde das Konzept für einen Identifikator CHE (Cultural Heritage Entities) entwickelt. Derzeit ist

geplant, dass die Deutsche Nationalbibliothek Betreiber des neuen PI ist, den Namensraum für CHE festlegt und den Resolving-Dienst bereitstellt. Partner in den jeweiligen Fachgemeinschaften sollen für das Vergabesystem und den direkten Kontakt mit den Anwendern zuständig sein. Zentraler Punkt der Zusammenarbeit ist der Austausch über die erforderlichen Metadaten. Es wird davon ausgegangen, dass jede Community eigene spezifische Bedarfe haben wird, die überein gebracht werden müssen. Eine erste prototypische Umsetzung ist mit dem Bundesarchiv geplant. Als nächsten Schritt wird mit Museen und Mediatheken über das Konzept und Metadaten-schemata diskutiert⁷.

Eindeutige Identifizierung von Autoren. Autorenidentifikatorsysteme wie ORCID sollen den Bezug zwischen einer Person und ihrem wissenschaftlichen Werk herstellen. Ziel ist die Reputationszuweisung. Zentral für die Akzeptanz solcher Systeme ist der Abbau von Zugangsbarrieren. Insbesondere Mehrfacheingaben werden von Wissenschaftlern als störend empfunden. ORCID ist kein fachspezifisches System und adressiert die Forschungen direkt. Wissenschaftliche Einrichtungen können ihre Autoren bei der Erlangung einer Autoren-ID unterstützen, ORCID sieht allerdings vor, dass die Seite vom Autor selbst gepflegt wird. Bislang ist ORCID kein Standard, wenngleich einzelne Einrichtungen wie das FZ Jülich begonnen haben, ihre Wissenschaftler anzuhalten, sich dort anzumelden. Im Falle einer institutsweiten Umsetzung sind Policies hilfreich. Allerdings erlauben viele institutionelle Systeme wie Repositorien und Forschungsinformationssysteme (FIS) keine Integration entsprechender Identifier. Derzeit wird der Haupttraffic im Web zu 80% durch Maschinen erzeugt. Diese können häufig PI nicht interpretieren, sondern folgen lediglich den Links⁸.

Hybrid Open Access und Double Dipping. Als Double Dipping wird die fehlende oder unzureichende Anpassung der Subskriptionspreise von Zeitschriften bezeichnet, die hybrides Open Access anbieten (d.h. Freischaltung von einzelnen Publikationen, für die Publikationsgebühren – APCs – bezahlt wurden). Für Inhalte aus diesen Zeitschriften wird somit doppelt bezahlt, einmal über die Subskriptionsgebühren und dann über die

Publikationsgebühren. In einer Befragung großer Wissenschaftsverlage nach ihrer No-Double-Dipping-Policy wurde festgestellt, dass lediglich 16 der untersuchten Verlage über eine entsprechende Policy verfügen. Verlage ohne Policy geben häufig an, dass die Anzahl ihrer Open-Access-Publikationen zu gering sei, um Preisanpassungen bei den Subskriptionspreisen vornehmen zu können. Bei Verlagen mit einer No-Double-Dipping-Policy lässt sich beobachten, dass diese die Preise häufig nicht in dem Maße reduzieren wie zu erwarten wäre:

- Eine Reduktion findet nur dann statt, wenn die Anzahl der Toll-Access-Publikationen sinkt.
- Subskriptionsgebühren werden erst dann angepasst, wenn ein bestimmter Schwellenwert erreicht ist.
- Lediglich der Listenpreis wird reduziert, d.h. bei Zeitschriftenpaketen schlägt sich die Anpassung nicht nieder.

Enttäuschend ist nach wie vor, dass die Verlage, die keine No-Double-Dipping-Policy betreiben, kaum Einblick in ihre Preisgestaltung gewähren. Erfolgsversprechend scheinen Modelle zu sein, nach denen die Einrichtungen die Summe an Subskriptionsgebühren teilweise zurückerstattet bekommen, die sie für Publikationsgebühren aufgewendet haben. Erste Pilotprojekte hierzu laufen in Großbritannien und Österreich. Die ursprüngliche Annahme, dass hybride Zeitschriften Mittel zur Transformation zu Open Access sein könnten, hat sich als Irrtum erwiesen. Neben Double Dipping sind die Gründe hierfür:

- Publikationsgebühren (APCs) bei hybriden Zeitschriften sind höher als bei reinen Open-Access-Zeitschriften;
- schlechtere Sichtbarkeit der Publikation, weil für Außenstehende oft nicht sofort ersichtlich ist, welche Publikationen freigeschaltet sind. Damit einher geht ein fehlender „Citation Advantage“, den Artikel aus reinen Open-Access-Zeitschriften auf sich verbuchen können;
- unbefriedigende Rechtsposition, da Autoren nicht immer ihre Rechte behalten, wie es bei Open-Access-Zeitschriften üblich ist;
- ablehnende Haltung der Forschungsorganisationen wie DFG gegenüber hybride Zeitschriften, demzufolge auch keine Förderung⁹.

Finanzierung von Artikelgebühren. Eine Möglichkeit, die Höhe der Publikationsgebühren zu kontrollieren, ist das Modell von SCOAP³ (Sponsoring Consortium for Open Access Publications in Particle Physics). In der Physik spielen Veröffentlichungen auf dem Goldenen Weg des Open Access und die Qualitätssicherung mittels Peer-Review-Verfahren eine wichtige Rolle. Das Konsortium mit Partnern weltweit hat sich zum Ziel gesetzt, die Publikationsgebühren durch ein Ausschreibungsverfahren zu deckeln. Ziel ist es, für die APCs einen Wettbewerb in Gang zu setzen und so für mehr Transparenz zu sorgen. Zudem wird versucht, eine Kappungsgrenze einzuführen. Hierzu werden die Subskriptionsgebühren, die bisher für die einschlägigen Zeitschriften bezahlt wurden, in einen Topf gegeben. TIB übernimmt für die Hochschulen in Deutschland die Vertretung im Konsortium. Deutschland ist mit 9,1% der zweitgrößte Mittelgeber. 50% der Hochschuleinrichtungen beteiligen sich an der Finanzierung, allerdings profitieren auch die übrigen Hochschulen von dem Modell, weil die finanzierten Publikationen im SCOAP³-Repositorium und auf den Verlagsseiten frei zugänglich sind. Bislang wurden etwa 5.800 Publikationen hierüber finanziert. Aktuell steht eine neue Ausschreibungsrunde an. Künftig sollen Institutionen nach ihrem Publikationsaufkommen in den Topf einzahlen. Ob SCOAP³ weiter parallel zum ArXiv bestehen kann, muss die wissenschaftliche Community entscheiden. Der Vorteil von SCOAP³ ist, dass die Publikationen im Gegensatz zur Veröffentlichung im ArXiv ein Peer-Review-Verfahren durchlaufen haben und die Publikationen mit der Veröffentlichung durch den Verlag frei verfügbar sind¹⁰.

Infrastrukturen für Open Access: Re3data.org. Die Landschaft der Forschungsdatenrepositorien ist heterogen. Es gibt fachliche Ansätze wie PANGAEA, institutionelle Repositorien wie an der LMU München, aber auch projektbezogene und generische Ansätze wie ZENODO und FIGSHARE. Der Dienst re3data.org will Forschungsdatenrepositorien verschiedener Disziplinen erfassen und soll Wissenschaftlern und anderen Akteuren im Wissenschaftssystem helfen, geeignete Repositorien zu finden. Hierzu wurde ein ausführliches Schema entwickelt, welches neben allgemeinen

Informationen auch Details zu Verantwortlichkeiten, Policies, Informationen zur Lizenzierung sowie technische und qualitative Standards sammelt. Repositoriumsbetreiber und andere Stellen können ein Repository zur Aufnahme in re3data.org vorschlagen. Die Angaben werden zunächst von einem Herausbergremium formal geprüft, bevor der Eintrag veröffentlicht wird. Bis Mitte April waren mehr als 1.200 Repositorien registriert, derzeit kommen pro Woche etwa zehn Einträge hinzu. Die Betreuung des Dienstes sowie die Finanzierung des technischen Betriebs werden künftig von DataCite übernommen, der langfristige Betrieb wird durch das KIT gesichert sein. Wichtigstes Prinzip der Plattform ist Offenheit. Es wird mit offenen Schnittstellen gearbeitet, auch das Metadatenschema ist offen publiziert, Metadaten und Website stehen unter einer CC-Lizenz und sind frei verfügbar¹¹.

COAR. COAR (Cooperation of Open Access Repositories) verfügt weltweit über hundert Mitglieder und betreibt die Vernetzung von Open-Access-Repositorien. Ziel ist die Schaffung einer „Sozialen Infrastruktur“ zum Austausch. Der Fokus liegt auf den grünen Weg des Open Access. Die Interoperabilität zwischen den Repositorien soll verbessert werden und so die Sichtbarkeit der Dokumente sowie der Repositorien selbst, in denen diese Dokumente abgelegt sind. Denn viele Publikationen, die in Repositorien liegen, werden über Suchmaschinen gefunden, ohne dass die Nutzer wissen, dass es das jeweilige Repository gibt¹².

e-infrastructures Austria. Ziel des dreijährigen Projekts, an dem zwanzig österreichische Universitäten sowie fünf außeruniversitäre Einrichtungen teilnehmen, ist die Schaffung und der Ausbau von Repositoriumsinfrastrukturen für Forschung und Lehre sowie die Etablierung eines effizienten Forschungsdatenmanagements. Mit ihm soll die Sichtbarkeit und Standardisierung der Repositorien vorangetrieben werden, ebenso die Harmonisierung der Policies. Dazu wird auf zwölf Themenfeldern zusammengearbeitet, die sich unter anderem auf rechtliche Fragen und das Data Life Cycle Management beziehen. Zudem soll es eine nationale Befragung zu Forschungsdaten geben¹³.

Publikationsdatenmanagement und Forschungsinformationssysteme. Publikationsdatenmanagement dient der Erfassung institutioneller Publikationen, die für die Forschungsevaluation, Drittmittelanträge und Berichterstattung wichtig sind. Sie können Grundlage für das Forschungsmarketing sein und helfen den Autoren bei der Erstellung von Publikationslisten. Insbesondere für die Ermittlung von Outputindikatoren ist unabdingbar,

dass die institutionelle Adresse einheitlich angegeben wird. Die Universität Nürnberg-Erlangen hat hierzu eine Broschüre mit Empfehlungen herausgegeben, die den eigenen Autoren Hilfestellungen gibt¹⁴. Zur Erstellung war notwendig, dass sich Einrichtungen wie Präsidium, Universitätsbibliothek, Marketingabteilung und Forschungsabteilung über Zielsetzungen und Ausgaben verständigten. Ziel ist insbesondere das Erreichen einer hohen Qualität der Metadaten¹⁵. Durch den Einsatz eines Forschungsinformationssystems (FIS) soll auch die Erfassung der Publikationen erleichtert werden. Die Einführung eines entsprechenden Systems und das Einpflegen rückwirkender Publikationsdaten ist alles andere als trivial. Zusätzlich zu einer Vereinheitlichung der Metadaten und zum Mapping dieser Daten auf das jeweilige System müssen Wissenschaftler an der Einrichtung motiviert werden, ihre Publikationsdaten zu melden. Idealerweise tragen Wissenschaftler ihre Publikationen eigenständig in das FIS ein und werden regelmäßig hieran erinnert¹⁶.

1 Ulrich Meyer-Dörpinghaus, Beate Tröger: *Forschungsdatenmanagement als Herausforderung für Universitäten und Universitätsbibliotheken.*

Jana Neumann, Frauke Ziedorn: *Data Librarians – Forschungsdatenmanagement als neue Aufgabe für Bibliotheken.*

Ania López: *Forschungsdatenmanagement an deutschen Universitäten. Ein neues Arbeitsfeld für Bibliotheken?*

2 Sven Vlaeminck: *Aktive Vernetzung in der Wissenschaftscommunity – Zwei Beispiele aus dem Bereich Forschungsdaten in den Wirtschaftswissenschaften.*

3 Ulrich Meyer-Dörpinghaus, Beate Tröger: *Forschungsdatenmanagement als Herausforderung für Universitäten und Universitätsbibliotheken.*

4 Ulrich Meyer-Dörpinghaus, Beate Tröger: *Forschungsdatenmanagement als Herausforderung für Universitäten und Universitätsbibliotheken.*

Jana Neumann, Frauke Ziedorn: *Data Librarians – Forschungsdatenmanagement als neue Aufgabe für Bibliotheken.*

Birte Lindstädt: *Forschungsdatenmanagement als überregionale Aufgabe der Informationsversorgung - was kann eine Zentrale Fachbibliothek wie ZB MED – Leibniz-Informationszentrum Lebenswissenschaften für andere Bibliotheken und die Wissenschaft leisten?*

Ania López: *Forschungsdatenmanagement an deutschen Universitäten. Ein neues Arbeitsfeld für Bibliotheken?*

5 Jana Neumann: *DataCite und DOIs für Forschungsdaten.*

6 Matthias Razum: *Persistent Identifiers für*

dynamische Inhalte – Anforderungen und Konzepte.

7 Uwe Müller: *Persistent Identifiers für Kulturobjekte – Fachkonzept für den PIDienst CHE.*

8 Najko Jahn: *Persistent Identifiers jenseits von Textpublikationen – Perspektiven und Entwicklungsbedarf.*

9 Bernhard Mittermaier: *Das „Double Dipping“ beim Hybrid Open Access. Detailliertere Ergebnisse lassen sich nachlesen in: Bernhard Mittermaier (2015): Double Dipping beim Hybrid Open Access – Chimäre oder Realität? In: Informationspraxis 1 (1), <http://dx.doi.org/10.11588/ip.2015.1.18274>.*

10 Angelika Kutz: *SCOAP3 – Start zum 01.01.2014: Gold Open Access in der Hochenergiephysik.*

11 Heinz Pampel, Frank Scholze: *Vom Projekt zum Service – Zur Rolle von re3data.org in der Forschungsdateninfrastruktur.*

12 Wolfgang Horstmann, Maxie Putlitz: *Confederation of Open Access Repositories: Weltweit aktiv zur Vernetzung von Open-Access-Repositorien.*

13 Paolo Budroni: *e-infrastructures Austria. Bibliothekarische Strategien zur Projektabwicklung auf nationaler Ebene.*

14 Leitfaden Affiliation: <https://www.fau.de/intranet/corporate-identity/leitfaden-affiliation-deutsch-und-englisch/> (zuletzt abgerufen am 29.06.2015).

15 Markus Putnings, Sabine Nunius: *Publikationsdatenmanagement als kollektive Herausforderung für Universitätsbibliothek, Sprachendienst und Forschungsmanagement: Lösungsansätze und Strategien.*

16 Gisela Weinerth, Jennifer Adler: *Integration der Universitätsbibliographie der Universität Hamburg in ein Forschungsinformationssystem (FIS).*

Die Präsentationen der hier genannten Beiträge sind teilweise abrufbar über BIB-OPUS unter: <https://opus4.kobv.de/opus4-bib-info/solrsearch/index/search/searchtype/collection/id/16253>.