

Forschungsdatenmanagement: Ein neues Berufsfeld?

Noch verschwinden viele Forschungsdaten und bleiben somit ungenutzt

Handbuch Forschungsdatenmanagement / Herausgegeben von Stephan Böttner, Hans-Christoph Hobohm; Lars Müller. Bad Honnef: Bock + Herchen, 2011. 223 Seiten: Illustrationen. 978-3-88347-283-6 – broschiert, 24,90 Euro. – Kostenlos erhältlich online unter: www.forschungsdatenmanagement.de/



Anschrift des Rezensenten: **Tillmann Tegeler**, Institut für Ost- und Südosteuropaforschung, Bibliothek und elektronische Forschungsinfrastruktur, Landshuter Str. 4, 93047 Regensburg, E-Mail: tegeler@ios-regensburg.de

In seinen »Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020«¹ vom 13. Juli 2012 verweist der Wissenschaftsrat auf Forschungsdaten als eines der zentralen Aufgabenfelder im Rahmen einer Gesamtstrategie für Informationsinfrastrukturen. Dabei kommt dem Umgang mit Forschungsdaten neben der Bereitstellung nicht-digitaler Medien, Digitalisierung und Langzeitarchivierung eine Schlüsselrolle beim wissenschaftlichen Arbeiten zu. Mit ihren Vorschlägen von 1998² hat die Deutsche Forschungsgesellschaft (DFG) die Sicherung und Nachnutzung von Forschungsdaten zu einem Bestandteil guter wissenschaftlicher Praxis erklärt. Seitdem ist eine Vielzahl an Datenrepositorien entstanden; Gremien und Kommissionen wurden ins Leben gerufen, um Strategien für das Forschungsdatenmanagement (FDM) zu formulieren.

Das im Folgenden besprochene Handbuch Forschungsdatenmanagement soll dabei als Leitfaden dienen, der über Institutionen und Entwicklungen informiert, die inzwischen oft an den Bedürfnissen der Fachdisziplinen ausgerichtet sind. Zu diesem Zwecke haben die Potsdamer Herausgeber zahlreiche Autoren gewonnen, die einzelne Aspekte dieses in den letzten Jahren von Wissenschaftsorganisationen und -förderern immer stärker in den Vordergrund der Diskussion um den Ausbau der Forschungsinfrastruktur gerückten Themas zu diskutieren.

Transparenz und Nutzerfreundlichkeit als Wegmarken

Das Buch gliedert sich in drei Teile: Auf einführende Betrachtungen folgen Praxisbeispiele, an die sich Konzepte zum FDM anschließen. Für Einsteiger, an die sich das Handbuch ebenso richtet wie an im FDM Beschäftigte, bieten die ersten Kapitel einen guten Überblick über Diskussionsstand und Terminologie. Dabei wird ebenso auf die beiden Grundaufgaben des FDM (Nachnutzung und Überprüfbarkeit) eingegangen, wie darauf verwiesen wird, dass das Thema »Qualitätssicherung« ein Desiderat im Handbuch darstellt. Letzteres ist bedauerlich, aber dem aktuellen Forschungsstand geschuldet. Für das vierte Forschungsparadigma, wie Jim Gray eine datengetriebene Wissenschaft 2007³ charakterisierte, müssen folgende Voraussetzungen geschaffen werden: Investitionen in Infrastruktur, Erstellen von Regelwerken sowie eine bessere personelle Ausstattung der Institutionen.

Gerade die Beschreibung und langfristige Interpretierbarkeit der Daten erfordern neue Berufsbilder, wie die Autoren einhellig anmerken. Noch ist der Datenkurator ein äußerst seltener informationswissenschaftlicher Zeitgenosse. Er muss den Wissenschaftler, der Daten erstellt und bearbeitet, im Forschungsprozess unterstützen und begleiten. Daten sollen nicht mehr im Papierkorb verschwinden oder auf privaten Festplatten ruhen, sondern durch »data sharing« einen neuen Lebenszyklus erfahren. Als Modelle seien das »Curation Lifecycle Model« oder das »Data Curation

Der Datenkurator ist noch ein äußerst seltener informationswissenschaftlicher Zeitgenosse.

Continuum« genannt. Dabei kommt der Selektion von speicherwürdigen Daten eine besondere Bedeutung zu.

Beim FDM selber stehen Transparenz und Nutzerfreundlichkeit im Vordergrund. Hierfür müssen Daten dokumentiert, Metadaten standardisiert und die langfristige Bereitstellung gewährleistet werden. Damit bei dem Erstellen und der Beschreibung von Daten Standards eingehalten werden, bedarf es einer Würdigung dieser Tätigkeit. Für Wissenschaftler sollte die Produktion von Daten deshalb als Forschungsleistung anerkannt werden. Trotz der in vielen Disziplinen erkennbaren Absicht zum »data sharing« ist die Umsetzung bei der Bereitstellung von Daten noch immer defizitär, was teilweise an fehlenden »Policies« liegt. Auch sind rechtliche Fragen vor allem zur Nutzung oftmals noch ungeklärt. Um viele Unklarheiten bereits im Vorfeld des FDM auszuräumen, stellt Uwe Jensen (GESIS) die Anforderungen an Datenmanagementpläne am Beispiel der Sozialwissenschaften dar. Ziel ist es, Daten »für Analysen und Replikationen dauerhaft und global verfügbar

- 1 www.wissenschaftsrat.de/download/archiv/2359-12.pdf [Zugriff am 28.11.2012]
- 2 www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf [Zugriff am 28.11.2012]
- 3 siehe http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_complete_lr.pdf, S. XIX-XXXIII [Zugriff am 28.11.2012]
- 4 www.pangaea.de/ [Zugriff am 28.11.2012]
- 5 www.dfg.de/formulare/54_01/54_01_de.pdf [Zugriff am 28.11.2012]
- 6 www.forschungsdatenmanagement.de [Zugriff am 28.11.2012]

[zu machen]« (Seite 71). Das umfasst die Dokumentation der Daten, die Erhebung umfassender Metadaten, Fragen der Organisation und Qualitätssicherung sowie Konzepte zur Langzeitarchivierung.

Funktionalität von Repositorien muss erweitert werden

Eine besondere Beachtung sollte dabei die Notwendigkeit zur interdisziplinären Nutzung der Daten erfahren, wobei die aus Gründen größtmöglicher Akzeptanz fachspezifisch normierten Standards eine besondere Herausforderung darstellen. Zur Verwaltung der Forschungsdaten werden Repositorien genutzt, an die zunehmend der Anspruch auf Interoperabilität gestellt wird, und deren Funktionalität damit weit über die reiner Publikationsrepositorien hinausgehen muss. Beispielhaft wird immer wieder die Datenbibliothek der Umweltwissenschaften, Pangaea,⁴ angeführt. Hieran zeigt sich exemplarisch, wie stark »Open Access« die kollaborative Arbeit beeinflussen kann. Doch muss immer auch auf die wissenschaftliche Kultur der Fachdisziplinen Rücksicht genommen werden, um eine möglichst hohe Akzeptanz zu erreichen – denn gerade in den Geistes- und Sozialwissenschaften gehört der OA-Gedanke aus unterschiedlichen Gründen nicht zur täglichen Praxis. Dies gilt es auch bei Strategien zur Langzeitarchivierung (LZA) zu berücksichtigen, wenn Zugriffe auf die Daten beschränkt werden. Grundsätzlich muss die Integration der LZA ebenso in die Überlegungen zu den Arbeitsabläufen während des Forschungsprozesses mit einfließen, wie

Daten sollen nicht mehr im Papierkorb verschwinden oder auf privaten Festplatten ruhen, sondern durch data sharing einen neuen Lebenszyklus erfahren.

Datenmanagementsysteme (DMS) ihre Anforderungen an die Erfassung, Anreicherung und Bereitstellung von Forschungsdaten ausrichten müssen. Man unterscheidet die meist genutzten »Collaboration Domains« von den eher spezialisierten »Publication Domains«. Um größtmögliches Vertrauen hervorzurufen, muss dabei Information über jeden Schritt im Rahmen des Datenmanagements erfasst und dokumentiert werden können. Dies umfasst die Verknüpfung der Daten mit Metadaten, die Versionierung sowie Datenformate, semantische Relationen zwi-

schen Datenobjekten, Persistenz und Authentifizierung. Abgelegt sind die Daten, ihre Beschreibungen und ihre Relationen in unterschiedlichen Schichten: der Persistenzschicht, dem eigentlichen »Lager« für die Daten, der Kernschicht mit den Metadaten und der Diensteschicht, die dem Nutzer unterschiedliche Handhabungen der Daten liefert.

Diese Architektur bietet die Möglichkeit zur langfristigen Archivierung der Daten oder zum verteilten Arbeiten, wie das »Grid-« oder »Cloud-«-Systeme erlauben. Schließlich sollen Repositorien auch Tools zur Datenanalyse und -visualisierung anbieten. Daten sollten auffindbar, recherchierbar und analysierbar sein, womit der Datenbeschreibung mittels Ontologien eine zentrale Rolle im FDM zukommt. Den Forderungen nicht nur nach Interpretierbarkeit der Daten, sondern auch deren Unversehrtheit und Lesbarkeit müssen künftig »Data Curation Services« nachkommen, wie dies beispielsweise für die Klimatologie vom »World Data Center for Climate« übernommen wird. Damit soll sichergestellt werden, dass die Voraussetzungen für eine Förderung erfüllt werden. So fordert die DFG seit 2010 die nachhaltige Sicherung von Daten zu deren Nachnutzung.⁵

Dokumentation und Archivierung bereits beim Forschungsprozess notwendig

Wem bei einer Publikation der Forschungsdaten letztlich die einzelnen Aufgaben für eine Bereitstellung zukommen, hängt vom Publikationsmodell ab; die Langzeitarchivierung kann bei Repositorien, aber auch bei Verlagen liegen. Wichtig bleibt: Zum Zwecke der Nachnutzung und Überprüfbarkeit müssen Forschungsdaten in einer vertrauenswürdigen Umgebung publiziert werden und interoperabel sein. In ihrem Beitrag zeigen Günther Neher und Bernd Ritschel die Möglichkeiten des »Semantic Web« für die Vernetzung von Forschungsdaten auf. Neben den Chancen weisen sie auch auf die Probleme hin, die durch die Inkompatibilität von relationalen Datenbank mit dem »Semantic Web« entstehen. Zu Recht erklären sie noch einen hohen Forschungsbedarf, aber ermutigen zu neuen Wegen. Noch besteht aber das größte Problem, dass ein Großteil der Daten gar nicht gespeichert wird. Während im 19. Jahrhundert Aufsätze oft mit Forschungsdaten im Anhang veröffentlicht worden sind (Seite 191), ging diese Publikationstechnik im 20. Jahrhundert verloren. Dafür wird vielfach ein fehlendes Anreizsystem verantwortlich

gemacht. Erst in den letzten Jahren wurde eine Förderung an eine Verpflichtung zur Archivierung gekoppelt; bleibt das Problem der Selektion speicherwürdiger Daten. Klar dagegen ist nach dem derzeitigen Stand, dass Dokumentation und Archivierung bereits während des Forschungsprozesses erfolgen müssen.

Es fehlt noch die fachwissenschaftliche Expertise

Abgeschlossen wird das vorliegende Werk durch einen Diskussionsbeitrag zur Zukunft des Informationswissenschaftlers im FDM. Die Herausgeber Büttner und Hobohm kommen unter Mitarbeit von Stefanie Rümpel zu der Feststellung, dass es kaum »Data Curators« oder »Data Librarians« in Deutschland gebe, dies aber künftige Betätigungsfelder von Informationswissenschaftlern und Bibliothekaren würden, die den Forschungsprozess begleiten und nicht erst nach der Publikation die Ergebnisse beschreiben. Jedoch fehle im Moment – anders als in den USA – diesen die (fach)wissenschaftliche Expertise. Auch hier gilt, dass die Übergänge zwischen Wissenschaft, Bibliothek und IT fließend werden.

Insgesamt stellt das besprochene Handbuch eine hervorragende Einführung in den Komplex des Forschungsdatenmanagements dar. Jedes Kapitel ist in sich geschlossen, weshalb es auch nicht weiter stört, dass immer wieder auf die gleichen Belegstellen verwiesen wird. Etwas störend ist dagegen, dass die Links der Online-Version wegen Zeilenumbruchs nicht einfach durch Anklicken funktionieren, sondern in den Browser kopiert werden müssen. Jedoch wird dieses Manko durch die OA-Stellung unter einer CC-BY-NC-SA-Lizenz wieder ausgeglichen.⁶

Tillmann Tegeler