

Zwischen „Long Tail“ und Langzeitarchivierung

Welchen Beitrag leisten Bibliotheken zum Datenmanagement?

| MATTHIAS TÖWE | **Der rasante Anstieg digitaler Forschungsdaten lässt die Frage nach effektiver und nachhaltiger Archivierung immer drängender werden. Wie kann Forschungsdatenmanagement und Langzeitarchivierung organisiert werden? Welche Rolle spielen dabei wissenschaftliche Bibliotheken und vor welchen Herausforderungen stehen sie?**

Die Aufgabe, eigene Daten zu organisieren, ist nicht neu für Forscher. Mit der Digitalisierung immer weiterer Prozesse und der zunehmenden Vielfalt und Menge von Forschungsdaten steigen die Anforderungen an das Datenmanagement allerdings. In den vergangenen Jahren ist es zudem von den Förderorganisationen verstärkt als eigenständiges Thema aufgegriffen worden, da das Datenmanagement auch die Voraussetzungen für eine spätere Nachnutzung von Daten schafft. Damit verbunden ist die Hoffnung, Forschungoutputs wie z.B. Software oder nachnutzbare Forschungsdaten aufzuwerten gegenüber der Fokussierung auf reine Publikationslisten. Aus Sicht der Hochschulen kommt hinzu, dass Mängel im Datenmanagement zu Reputationsrisiken für Forscher und ihre Hochschule führen können.

Das Datenmanagement ist nicht etwa nur eine Herausforderung für Disziplinen, die große Datenmengen nutzen. Die Anforderungen an Workflows und Technik sind in diesen Fächern

zwar höher, der Bedarf nach grundlegenden Konzepten und Maßnahmen zur Datenorganisation besteht jedoch auch bei eigentlich überschaubaren Datensammlungen des sogenannten „Long Tail“. Welchen Beitrag leisten Hochschulbibliotheken zum Forschungsdatenmanagement bereits bzw. worin könnten sie die Forscher zusätzlich unterstützen?

»Im Extremfall hat aber nur genau eine Person die vollständige Orientierung über die Datensammlung.«

Brauchen Forschende überhaupt Unterstützung?

Forscher kennen ihre eigenen Forschungsdaten am besten. Im Extremfall hat aber nur genau eine Person aufgrund ihres impliziten Wissens die vollständige Orientierung über die Datensammlung. Dies birgt für die Auffindbarkeit und langfristige Nutzbarkeit von Forschungsdaten erhebliche Risiken, zumal die Mobilität von Forschenden je nach Phase ihrer Laufbahn hoch ist.

In Disziplinen mit einer Tradition des Datenaustauschs wurden früh Anstrengungen unternommen, um diese Risiken zu mildern und als wertvoll oder unersetzlich erkannte Datenbestände zu dokumentieren und zu sichern. Anforderungen von Forschungsförderern und Institutionen zielen darauf ab, dies in der Breite der Diszipli-

nen zu erreichen. Zwar muss letzten Endes jede Forschungsgruppe ihren eigenen angemessenen Umgang mit den Anforderungen und ihrer praktischen Umsetzung finden; es ist aber nicht effizient, wenn jede Gruppe dies von Grund auf neu angeht. Innerhalb einer Hochschule ist es darum sinnvoll, Know-how zu den Anforderungen und zu möglichen Umsetzungen in der Praxis zu bündeln und in Kursen und Beratungen weiterzuvermitteln.

Datenmanagementplanung und Beratungsbedarf

Ein Beispiel dafür sind Anforderungen von Geldgebern – in unserem Fall des Schweizerischen Nationalfonds (www.snf.ch) – zur Einreichung von Datenmanagementplänen (DMPs) mit Projektanträgen. Die Reaktionen von Forschenden dazu reichen vom ernsthaften Bemühen, die als sinnvoll erachtete Anforderung zu erfüllen, über die Befürchtung einer weiteren Bürokratisierung bis hin zur Bewertung als reine „box-ticking exercise“. In jedem Fall werfen die Anforderungen wiederkehrende Fragen auf, zu denen Informationen von zentralen Einrichtungen der Hochschule gesammelt und bereitgestellt werden können. Die niederschwellige Erreichbarkeit vor Ort hat dabei eine hohe Bedeutung.

Für eine Ansiedlung dieser Aufgabe bei Bibliotheken sprechen Berührungspunkte zu Themen wie dem Publizieren unter Open Access, der zitierfähigen Bereitstellung von Daten mit persistenten Identifikatoren wie dem Digital Object Identifier (DOI) oder der Verwendung von eindeutigen Autorenidentifikationen wie der Open Researcher and Contributor ID (ORCID). Diese Dienste werden von Hochschul-

AUTOR



Dr. **Matthias Töwe** ist Leiter Forschungsdatenmanagement und Datenerhalt an der ETH-Bibliothek Zürich.

bibliotheken technisch sowie durch den Abschluss von Vereinbarungen mit den Anbietern ermöglicht und mit Beratungs- und Kursangeboten unterstützt. In der Praxis führen viele Gespräche zum Datenmanagement überraschend schnell auf diese und verwandte Themen, so dass eine Zusammenfassung bei der Bibliothek – unter Einbezug der jeweiligen Akteure vor Ort – in vielen Fällen einleuchtet.

Bereitstellung von Daten nach den FAIR-Prinzipien

Seit einigen Jahren wird die pauschale Forderung nach Open Data durch die Formulierung besser fassbarer Erwartungen an die Bereitstellung von Forschungsdaten in Form der FAIR-Prinzipien konkretisiert (Findable, Accessible, Interoperable, Reusable mit jeweils mehreren Unterpunkten). Etliche Unterpunkte der FAIR-Prinzipien stellen letztlich Anforderungen an die Qualität und den Umfang der Metadaten dar. Daraus ergibt sich ein Bedarf, die teilweise unscharfen Prinzipien in reale Anwendungen und in konkrete, geeignete Metadaten zu übersetzen.

Bibliotheken folgen seit Jahrzehnten internationalen Metadatenstandards und tragen neue mit, die explizit für Forschungsdaten gedacht sind. Am verbreitetsten dürften derzeit das Metadaten-Schema des DataCite-Konsortiums sowie Varianten des Dublin Core Metadaten Schemas sein. Die Kompetenz, auf dieser Basis Schemata zu pflegen, die den FAIR-Prinzipien entsprechen, kann dort, wo sie zum großen Teil bereits existiert, gebündelt und ausgebaut werden.

Eine wesentliche Rolle spielen Metadaten beim Aufbau von Forschungsdatenrepositorien an Hochschulen, an denen Bibliotheken in der Regel maßgeblich beteiligt sind. Oft dürfte das eigene Repositorium als subsidiäre Ebene neben internationalen fachspezifischen und anderen externen Repositorien genutzt werden. So können Daten aus Fächern ohne anerkannte fachspezifische Datenbanken publiziert und archiviert werden, oder es ist ohnehin eine – allenfalls zusätzliche – Speicherung im Zugriff der eigenen Hochschule gewünscht. Auch ist zumindest ein Nachweis der an anderen Orten abgelegten Forschungsdaten in-

nerhalb der eigenen Institution wünschenswert. Diese Sichtbarkeit ist eine Voraussetzung dafür, dass die Bereitstellung hochwertiger Forschungsdaten stärker als eigenständige wissenschaftli-

»Bibliotheken folgen seit Jahrzehnten internationalen Metadatenstandards und tragen neue mit.«

che Leistung honoriert werden kann und ein Anreiz geschaffen wird.

Langzeitarchivierung

Der zunehmende Druck, Forschungsdaten für Dritte zugänglich zu machen, erhöht den Bedarf, ihre langfristige Verfügbarkeit zu erhalten. Damit ist wie erwähnt nicht die reine Speicherung der Daten gemeint, sondern im Idealfall darüber hinaus die Bewahrung des wissenschaftlichen Kontexts der Daten und weiterer Informationen und Objekte, die nach heutigem Wissen für eine spätere Nutzung notwendig sein werden. Dagegen dürfte eine klassische Formatmigration, also die Konvertierung eines obsoleten Dateiformats in ein neueres, nur bei einem Teil der Forschungsdaten eine größere Rolle spielen.

Für die notwendigen Prozesse können Bibliotheken Unterstützung bieten, insbesondere, wenn sie über Erfahrungen aus der Langzeitarchivierung ihrer

»Es ist nicht das Ziel, möglichst viele Forschungsdaten langfristig zu archivieren, sondern die jeweils relevanten Daten für angemessene Zeit zu erhalten.«

eigenen Daten verfügen. Dies gilt bereits bei der Organisation, Dokumentation und Auswahl von Daten. Weitergehende Leistungen können darin bestehen, Daten über Jahrzehnte hinweg, d.h. auch über die Lebensdauer von Forschungsgruppen hinaus, vorzuhalten und gegebenenfalls die Prozesse zu steuern, um Daten kontrolliert löschen oder ihre Aufbewahrungsdauer verlängern zu können.

Es ist nicht das Ziel, möglichst viele Forschungsdaten langfristig zu archivieren, sondern die jeweils relevanten Daten für angemessene Zeit zu erhalten.

Die Forschenden selbst können und müssen basierend auf den Gepflogenheiten und Kriterien ihres Fachs eine sinnvolle Auswahl treffen. Sie stehen dabei unter erheblichem Druck sachfremder Kriterien wie der langfristigen Kosten für die reine Datenspeicherung. Mit den Erwartungen von Forschungsförderern werden diese Kosten weiter steigen, ohne dass im gleichen Maß die Möglichkeit besteht, sie bei

den Förderern geltend zu machen. Die Hochschulen in die Pflicht zu nehmen, Speicherkapazitäten als Teil ihrer Grundausrüstung bereitzustellen, verlagert das Problem selbstverständlich nur.

Gestaltung und Anforderungen

Für Investitionen in das Forschungsdatenmanagement durch die verschiedenen Beteiligten gibt es gute Gründe. Es sollte allerdings vermieden werden, dass das Thema zum Selbstzweck wird oder – im anderen Extrem – als reine Pflichtübung abgetan wird. Vielmehr bleibt es ein Mittel zum Zweck im Dienste der wissenschaftlichen Qualität und muss als solches auch entscheidend von den Forschenden ausgestaltet werden.

Das Datenmanagement stellt hohe Anforderungen. Um den Ressourcenaufwand für die einzelnen Forschenden zu minimieren, kann die kontinuierliche Unterstützung durch zentrale Einrichtungen wie die Hochschulbibliotheken im Verbund mit den jeweiligen Rechenzentren gewährleistet werden. Die eigenen, fachlich fundierten Entscheidungen der Forschenden zum Umgang mit Forschungsdaten können die Institutionen nicht ersetzen, den Aufwand für unterstützende Services oder gar den Betrieb der notwendigen Infrastrukturen können sie dagegen zu einem nennenswerten Teil übernehmen. Dabei ist unbestritten, dass alle Institutionen und ihr Personal gemeinsam mit den Forschenden einen Lernprozess durchlaufen, der von verfügbaren Ressourcen, Erfahrungen und weiteren Gegebenheiten vor Ort abhängt.

Das Datenmanagement stellt hohe Anforderungen. Um den Ressourcenaufwand für die einzelnen Forschenden zu minimieren, kann die kontinuierliche Unterstützung durch zentrale Einrichtungen wie die Hochschulbibliotheken im Verbund mit den jeweiligen Rechenzentren gewährleistet werden. Die eigenen, fachlich fundierten Entscheidungen der Forschenden zum Umgang mit Forschungsdaten können die Institutionen nicht ersetzen, den Aufwand für unterstützende Services oder gar den Betrieb der notwendigen Infrastrukturen können sie dagegen zu einem nennenswerten Teil übernehmen. Dabei ist unbestritten, dass alle Institutionen und ihr Personal gemeinsam mit den Forschenden einen Lernprozess durchlaufen, der von verfügbaren Ressourcen, Erfahrungen und weiteren Gegebenheiten vor Ort abhängt.

Lesen Sie online: Was bedeutet es in der Praxis, einen Datenmanagement-Plan zu erstellen? Im Interview mit Forschung & Lehre spricht Juniorprofessorin Birgit Glorius von der TU Chemnitz über ihre Erfahrungen aus einem EU-geförderten Horizon-2020-Projekt (<http://t1p.de/37k7>).