



UZH News

16.04.2015

Forschungsdaten

Klarschiff im Datenmeer

Hinter dem sperrigen Begriff «Forschungsdaten-Management» versteckt sich ein wichtiger Teil der Forschungsarbeit: das Verarbeiten, Speichern, Teilen und Publizieren von Forschungsdaten. An der UZH will eine neue Arbeitsgruppe Wissenschaftlerinnen und Wissenschaftler dabei unterstützen.

Marita Fuchs



Forschungsdaten sind ihr Kerngebiet: Christian Fuhrer, Leitung Open Access an der Hauptbibliothek, Andrea Malits von der Zentralbibliothek und Peter Kunszt von der Zentralen Informatik der UZH. (v.l.n.r.) Bild: Frank Brüderli

Forschungsdaten – sie sind der Treibstoff der Hochschulen. Der online erhältliche wissenschaftliche Ausstoss vermehrt sich so schnell wie nie: Alle neun Jahre verdoppelt er sich zurzeit. Angesichts dieser Datenmenge wünschen sich viele Forschende eine langfristige und sorgfältige Archivierung ausgewählter Forschungsdaten, wie die Konferenz der Universitätsbibliotheken der Schweiz in einer Umfrage unter Schweizer Universitäten bereits vor ein paar Jahren festgestellt hat.

Doch wie dabei vorgehen? Im Moment liegt es meist bei den Forschungsgruppen selber, wie sie ihre Daten archivieren. Sie entscheiden, welche Daten in den Papierkorb wandern oder welche aufbewahrt werden. Eine definierte Vorgehensweise gibt es nicht. Das hat viele Nachteile.

Konsistente Vorgehensweise

Nehmen wir den Fall eines kurz vor der Emeritierung stehenden Professors. Er ist Geisteswissenschaftler, hat mit seinem Team viele Quellen ausgewertet, gleichzeitig auch selbst Daten erhoben und dokumentiert. Teilweise sind diese Daten für wissenschaftliche Publikationen verwendet worden, doch sein Material geht weit darüber hinaus und könnte für neue Studien genutzt werden. Abgelegt sind alle Daten auf einem Institutsserver. Die Verantwortung für regelmässige Backups liegt beim Informatikverantwortlichen des Instituts oder bei den Mitarbeitern des Professors.

Was passiert, wenn der Professor nicht mehr an der Universität arbeitet? Wie lange werden die Daten im Backup aufbewahrt, und wer kann und darf auf sie zugreifen? Gibt es ein Konzept, wie in diesen oder ähnlich gelagerten Fällen vorgegangen wird?

Grundlagen erarbeiten

All diese Fragen will eine Arbeitsgruppe aus Vertretern der Zentralbibliothek Zürich und der Hauptbibliothek sowie der Zentralen Informatik der UZH anpacken und dabei die Grundlage für ein umfassendes Forschungsdatenmanagement an der UZH erarbeiten. Federführend sind Andrea Malits von der Zentralbibliothek, Christian Fuhrer, Leitung Open Access an der Hauptbibliothek, sowie Peter Kunszt von der Zentralen Informatik der UZH. Sie arbeiten im Auftrag des Prorektorats Rechts- und Wirtschaftswissenschaften und der Zentral- sowie der Hauptbibliothek der UZH. Gemeinsam wollen sie einen Vorschlag erarbeiten für eine konsistente Vorgehensweise bei der Verarbeitung, Archivierung und Publikation von Forschungsdaten.

Daten für den Reviewprozess

Viele europäische Hochschulen arbeiten aktuell an neuen Archivierungssystemen. Handlungsbedarf besteht auch deshalb, weil immer mehr wissenschaftliche Journale Datenanhänge («supplemental data») oder ganze Datensets nicht mehr selbst publizieren wollen. Sie erwarten von den Forschenden, dass sie ihre Daten in so genannten Repositories zugänglich machen, also in vertrauenswürdigen und nachhaltigen Datenservern. «Die Journals wollen auch für den Reviewprozess zunehmend Zugriff auf die Originaldaten, und zwar in aufbereiteter Form», sagt Christian Fuhrer von der Hauptbibliothek der Universität Zürich. So etwa bei der Hirnforschung: Fallen viele Originalbilder von Hirnschnitten an, sollten sie auf einem Repository hinterlegt und einsehbar sein. Dazu müssen sie jedoch auch nach bestimmten Kriterien geordnet werden.

Fördergelder nur bei ausgeklügeltem Datenmanagement

Doch nicht nur die Journals, auch Förderungsinstitutionen sind zunehmend daran interessiert, dass bei Forschungsprojekten die Datenspeicherung gesichert ist. So etwa beim EU-Flaggschiff «Horizon 2020»: Reichen Forscher, deren Fachgebiet dort unter den «Pilot on Open Research Data» fällt, ihre Anträge ein, müssen sie angeben, wie viele Daten in etwa anfallen, wie sie zwischengespeichert und wo sie schlussendlich archiviert und zugänglich gemacht werden. Das kann mitentscheidend sein, ob ein Antrag akzeptiert wird oder nicht. Auch der Schweizerische Nationalfonds arbeitet im Moment an einem neuen Forderungskatalog zur Datenspeicherung.

Dass Forschungsdaten unter Berücksichtigung von Anforderungen des Datenschutzes, des geistigen Eigentums sowie von Konkurrenzsituationen archiviert und publiziert werden müssen, das verlangt eigentlich schon die gute wissenschaftliche Praxis, denn schliesslich sollen Experimente und Überlegungen überprüf- und nachvollziehbar sein. «Heute denken allerdings nur wenige Forschende an Archivierungszeiträume von mehr als zehn Jahren», sagt Fuhrer. Es fehlen auch einheitliche Vorgaben zur strukturierten Datenablage.

Metadaten geben Struktur

Forschungsdaten sind vielfältig und je nach Fachgebiet sehr unterschiedlich. Grob lassen sie sich jedoch in primäre und sekundäre Daten einteilen. Primärdaten, die etwa beim Experimentieren, Messen, aus Quellenmaterial oder Umfragen entstehen, werden in unterschiedlichen Arten aufbereitet. Es können digitale Tabellen, Bilder, Filme, Ergebnisse einer Umfrage, oder maschinell errechnete Resultate sein. Bei der Analyse dieser Primärdaten entstehen nach deren Auswertung so genannte Sekundärdaten.

Besonders wichtig für eine spätere Archivierung ist die Beschreibung aller Datensätze mit Metadaten, die alle Daten nach bestimmten Kriterien strukturieren. Denn erst die Metadaten erlauben ein schnelles Wiederauffinden. Dazu kommt die Formatauswahl und – besonders wichtig bei medizinischen Studien – die Anonymisierung personenbezogener Daten sowie die Beschreibung der experimentellen Protokolle und Workflows, mit denen die Daten erzeugt und verwertet wurden. Nur so können sich andere Forschende im Wust der Datensätze auch zurechtfinden und diese reproduzieren.

Unsichere Privatrechner

Das gilt für alle Fachbereiche, auch in den Geisteswissenschaften, die häufig mit digitalisierten Quellen arbeiten. «Wir

möchten diese so genannten Digitalisate und die digitale Arbeitsumgebung der Forschenden gezielt, das heisst nach einem Berechtigungssystem, zur Verfügung stellen können», sagt Andrea Malits von der Zentralbibliothek Zürich. Ein Beispiel: Arbeitet ein Forscher oder eine Forscherin zu den Schriften Pestalozzis, so sollten seine digitalisierten Quellentexte auch anderen Forschern zugänglich gemacht werden können, sofern dies gewünscht wird. Dafür braucht es eine geeignete IT-Infrastruktur. Liegen die Digitalisate jedoch lediglich auf einem Institutsrechner oder sogar auf dem Privatrechner, sind sie nicht greifbar und es besteht zudem die Gefahr, dass sie verloren gehen.

Doch auch wenn sich ein Forschungsteam entscheidet, Daten zu archivieren, stellen sich mehrere Fragen. Was muss archiviert werden? Und wie müssen die Daten gespeichert werden, damit auch andere darauf zugreifen können? Welche Daten sollen offen publiziert werden (Open Access zu Forschungsdaten, Open Data)? Datenmanagementpläne sollen Forschende anregen, sich mit diesen Fragen auseinanderzusetzen. «Wenn es klar ist, dass bestimmte Daten gespeichert werden sollen, dann müssen diese Daten gewisse Eigenschaften erfüllen, damit sie auch später von anderen gefunden und sinnvoll genutzt werden können», sagt Malits.

Unterstützung bei Datenmanagementplänen

Bei dieser Aufgabe will die Gruppe aus Andrea Malits, Christian Fuhrer und Peter Kunszt die Forschenden unterstützen. Ihr Ziel ist es, ab 2016 eine Service-Infrastruktur für Forschende aufzubauen. «Wir helfen bei der Erstellung von Datenmanagementplänen», sagt Peter Kunszt. Dabei werden vor Beginn der Forschungsarbeiten Schlüsselprobleme identifiziert, die bei der Arbeit mit den Daten entstehen können.

Auch sollen die Forschenden während des gesamten Forschungsprozesses unterstützt werden. «Denn manchmal gibt es neue Entwicklungen oder Wendepunkte in der Forschung, denen man gerecht werden muss», weiss Kunszt aus Erfahrung. Für die Erstspeicherung grosser Datensätze ist die UZH gerüstet, sagt der Informatiker. Seine Abteilung wurde vor einem Jahr eingerichtet, und er stellt die Soft- und Hardware zur Verfügung.

Ungelöste Fragen

Die bisherigen Kontakte der Arbeitsgruppe mit Forschenden der UZH haben gezeigt, dass einige Kernfragen des Umgangs mit Forschungsdaten an der Universität Zürich ungelöst sind: Offen ist vor allem die langfristige Archivierung von Forschungsdaten, aber auch die Veröffentlichung von Daten, wenn dazu kein geeignetes Fachrepository existiert, sowie die nachhaltige Pflege von Datenbanken, welche Forschende im Zug ihrer Projekte selber angelegt haben. Insgesamt resultiere aus Datenmanagement und insbesondere Archivierungsvorgaben für die Forschenden zunächst etwas mehr Aufwand, aber mit der Zeit werde sich ein greifbarer Mehrwert einstellen, ist Andrea Malits überzeugt.

Marita Fuchs, Redaktorin UZH News.

Links

Peter Kunszt, Zentrale Informatik UZH
Prof. Dr. Christian Fuhrer, Leiter Open Access ZORA, Hauptbibliothek UZH
Andrea Malits, Zentralbibliothek Zürich

Weitere Artikel

Die UZH gibt Schub bei der Datenanalyse, UZH News, 9.5.2014

top

