Forschungsdatenmanagement an der ETH-Bibliothek

Book Chapter**Author(s):**

Töwe, Matthias

Publication date:

2018

Permanent link:

<https://doi.org/10.3929/ethz-b-000283804>

Rights / license:

Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

Originally published in:

<https://doi.org/10.1515/9783110553796-015>

Matthias Töwe

Forschungsdatenmanagement an der ETH-Bibliothek

Abstract: Die ETH-Bibliothek beschäftigte sich als eine der ersten Hochschulbibliotheken im deutschsprachigen Raum mit dem Thema Forschungsdaten. Der Beitrag legt dar, wie der schrittweise Aufbau von Lösungen für verschiedene Aspekte des Forschungsdatenmanagements erfolgte und welche Infrastrukturen und Dienstleistungen heute angeboten werden. ‚Forschungsdatenmanagement‘ bezeichnet dabei nicht eine einzige, klar umrissene Aufgabe. Vielmehr handelt es sich um eine ganze Reihe von Anforderungen in verschiedenen Etappen des Lebenszyklus von Forschungsdaten – von der geordneten lokalen Ablage über die Bearbeitung und Auswertung bis hin zur Veröffentlichung und Langzeitarchivierung. Die einzelnen Phasen benötigen differenzierte Lösungen, die auch nicht allein von einer Bibliothek bereitgestellt werden können. Der Beitrag beleuchtet daher als Momentaufnahme auch den aktuellen Stand der Aufgabenverteilung innerhalb der ETH Zürich.

Einführung

Die Unterstützung des Forschungsdatenmanagements im engeren Sinn ist eine der jüngeren Aufgaben von Wissenschaftlichen Bibliotheken. Selbstverständlich sind auch früher schon Forschungsdaten als Teil von Nach- und Vorlässen oder in anderer Form in Bibliotheken und Archive gelangt. Mehr noch dienen und dienen ihre Bestände als Grundlage für Forschungsarbeiten, die sich keineswegs auf prominente Editionsprojekte beschränken.

Der folgende Beitrag betrachtet dagegen nur jenes Verständnis von Forschungsdatenmanagement, das sich in der Breite erst etwa seit 2010 etabliert hat und seinen Schwerpunkt auf digitale Daten legt. Trotz dieses überschaubaren Zeithorizonts wird der Versuch unternommen, am Beispiel der Dienstleistungen der ETH-Bibliothek verschiedene Etappen der natürlich keineswegs abgeschlossenen Entwicklung des Engagements von Bibliotheken beim Thema Forschungsdatenmanagement aufzuzeigen.

Die ETH-Bibliothek

Die ETH-Bibliothek wurde 1855 zusammen mit der ETH Zürich gegründet. Sie ist die zentrale Hochschulbibliothek für die ETH Zürich und wird als Schweizer Informationszentrum für Technik und Naturwissenschaften auch von einer breiten Öffentlichkeit genutzt. Wie in jeder Hochschule gilt das besondere Interesse der Kundinnen und Kunden zunächst einem umfassenden, aktuellen Angebot an Informationsquellen. Die Forderung nach einer umfassenden Versorgung beinhaltet zusätzlich bereits die Anforderung, einmal erworbene Inhalte auch auf Dauer zur Verfügung stellen zu können, und zwar auch in Fällen, in denen ein expliziter Archivauftrag allenfalls gar nicht besteht.

Diesen Bedürfnissen folgend hat die ETH-Bibliothek über Jahrzehnte einen Bestand mit grosser Breite und Tiefe zu den Fachgebieten aufgebaut, zu denen an der ETH Zürich geforscht und gelehrt wird. Der technischen Entwicklung folgend, sind auf diese Weise jeweils die zu ihrer Zeit aktuellen Medientypen in den Bestand gelangt. Dazu gehörten in erstaunlicher Zahl z.B. auch Mikrofichen und -filme, später Datenbanken zur Auftragsrecherche lediglich durch Fachpersonal und selbstverständlich seit etwa Mitte der 90er Jahre digitale Objekte aller Art. Deren Siegeszug war aufgrund der unverkennbaren Vorteile vielleicht theoretisch absehbar, die Geschwindigkeit des Wandels hin zu einer von elektronischen Dokumenten dominierten Informationsversorgung dürfte aber dennoch viele überrascht haben. Nicht zuletzt diese Geschwindigkeit führt dazu, dass bereits mehrere Generationen von Studierenden kaum noch einen Gedanken darauf verwenden, ob es neben den online verfügbaren Informationsressourcen allenfalls auch andere Bestände gibt, die für ihre Arbeit von Interesse sein könnten. Aktualität und Breite der digitalen Informationsversorgung verengen hier eher den Blick.

Überlegungen zur Langzeitarchivierung

Die Bestandserhaltung muss die Entwicklung zur digitalen Informationsversorgung nachvollziehen. Während zunächst die Sorge im Vordergrund stand, den Zugriff auf lediglich lizenzierte und nicht gekaufte Inhalte dauerhaft sicherzustellen, kamen bald auch technische Fragen hinzu. Wie lange bleiben aktuelle Dateiformate nutzbar, wer sorgt nicht nur für die reine Datenablage, sondern auch für die Auffindbarkeit der Inhalte?

An der ETH-Bibliothek wurden seit etwa 2001 verschiedene Projekte und Studien durchgeführt, teilweise unter Federführung des Konsortiums der Schwei-

zer Hochschulbibliotheken¹ sowie als Teil des Programms e-lib.ch: Elektronische Bibliothek Schweiz². Dabei wurden die Herausforderungen bereits recht klar benannt, es wurden aber auch Stolpersteine erkannt, die einer ‚einfachen‘ Lösung entgegenstanden. Aus diesem Grund verfolgte die ETH-Bibliothek die Frage der Langzeitarchivierung digitaler Ressourcen zwar kontinuierlich und mit einigem Aufwand weiter, konkrete Lösungen wurden aber nicht in Angriff genommen. Dies stand im Einklang mit der Strategie der ETH-Bibliothek, eher ‚early adopter‘ sein zu wollen als Pionier, weil Ressourcen für umfangreiche Eigenentwicklungen nicht zur Verfügung stehen.

Forschungsdaten im Blick

Die Auseinandersetzung mit der weiteren Entwicklung führte zunehmend zu der Erkenntnis, dass eine Fokussierung auf Informationsressourcen der Bibliothek zu kurz griff. Die Digitalisierung hatte selbstverständlich spätestens in den 2000er Jahren die eigentliche Forschung vollständig durchdrungen und speziell an einer technisch-naturwissenschaftlichen Hochschule waren viele Prozesse nur noch in digitaler oder zumindest in digital unterstützter Form denkbar. Während die fachwissenschaftlichen Aspekte der jeweiligen Abläufe naturgemäss von den Forschenden selbst beherrscht wurden, zeichnete sich ab, dass diese Entwicklung auf anderen Ebenen zunehmend Fragen aufwarf. So erzeugte die einfache Veränderbarkeit digitaler Daten durchaus Unbehagen, z.B. hinsichtlich der Frage, wie die Authentizität von Forschungsausputs gewährleistet werden könnte.

Demgegenüber stand und steht die Möglichkeit, erstmals auf vergleichsweise einfache Art Daten austauschen, überprüfen oder gar für neue Fragestellungen nachnutzen zu können. Dies führte schnell zu der Einsicht, dass das Potenzial für eine spätere Nutzung von Daten weitgehend davon abhängen würde, in welchem Umfang und mit welcher Qualität bereits zu einem frühen Zeitpunkt im Lebenszyklus Herkunft, Eigenschaften und Abhängigkeiten von Daten dokumentiert werden. Beide Gesichtspunkte sind eng miteinander verwoben: Daten können überhaupt nur dann seriös und mit Vertrauen nachgenutzt werden, wenn ihre Authentizität unstrittig ist. Umgekehrt ermöglichen es der Austausch oder gar die Publikation von Daten, die publizierten Ergebnisse im Detail nachzuvollziehen und eventuelle Unstimmigkeiten zu erkennen. Die Voraussetzungen für die verschiedenen Nutzungen müssen in jedem Fall früh durch ein konsequentes For-

1 <http://www.consortium.ch/>. Alle Links in diesem Beitrag wurden am 26.03.2018 überprüft.

2 <http://www.e-lib.ch/copy/de/Ueber-uns/Projekte/Langzeitarchivierung.html>.

schungsdatenmanagement geschaffen werden. Dementsprechend umfassend wurde auch der Auftrag der Fachstelle Digitaler Datenerhalt³ gefasst, die Ende 2010 zunächst als Projekt ihre Arbeit aufnahm.

ETH Data Archive

Eine umfassende Umfrage innerhalb der ETH Zürich zum aktuellen Umgang mit Forschungsdaten sowie zum Bedarf nach Unterstützung dafür lieferte 2011 wichtige Einblicke in die – sehr heterogenen – Bedürfnisse. Da die ETH-Bibliothek sich am intensivsten mit dem Teilaspekt der Langzeitarchivierung digitaler Daten befasst hatte, setzte sie auch im Hinblick auf das Forschungsdatenmanagement zunächst an diesem Punkt an. Der Aufbau des eigenen ETH Data Archive mit der seinerzeit noch recht jungen Applikation Rosetta⁴ der Firma Ex Libris ab 2011 wurde wesentlich getrieben durch das Ziel, Forschungsdaten dauerhaft verfügbar zu halten. Es war allerdings bereits zu diesem Zeitpunkt klar, dass damit keine umfassende Lösung für die Bedürfnisse der Forschenden angeboten werden konnte. Der grosse Bereich der Unterstützung des ‚Active Data Managements‘, also des Datenmanagements während der laufenden Forschungsarbeit mit kontinuierlicher Nutzung und Bearbeitung von Daten, wurde bewusst nicht adressiert. Zudem war klar, dass ein Langzeitarchiv prinzipiell kein geeigneter Aufbewahrungsort für grosse, unstrukturierte Datenvolumen („Big Data“) sein kann, deren Hauptzweck darin besteht, regelmässig mit datenwissenschaftlichen Methoden analysiert und prozessiert zu werden.

Forschungsdaten wurden somit eher als zentraler Bestandteil des wissenschaftlichen und kulturellen Erbes der ETH Zürich verstanden im Unterschied zu lediglich lizenzierten Inhalten von Dritten, die gleichzeitig an zahlreichen anderen Orten verfügbar sind. Für letztere Daten wurde daher eher eine Beteiligung an bestehenden Diensten wie LOCKSS⁵ und Portico⁶ ins Auge gefasst.

In dieser Phase zählten sich die zuvor gemeinsam mit den zentralen Informationsdiensten und dem zur ETH-Bibliothek gehörenden Hochschularchiv der ETH Zürich unternommenen Vorprojekte besonders aus: Sie hatten ein gemeinsames Verständnis der unterschiedlichen Aufgaben der Partner etabliert, das in der Folge die Zusammenarbeit und auch die Abgrenzung der Ansprechstellen aus Kundensicht wesentlich erleichterte. Das ETH Data Archive wurde vollständig in

3 <http://www.library.ethz.ch/Digitaler-Datenerhalt>.

4 <http://www.exlibrisgroup.com/category/RosettaOverview>.

5 <https://www.lockss.org>.

6 <http://www.portico.org>.

die Infrastruktur der Informatikdienste implementiert, während die ETH-Bibliothek mit eigenem Personal den Betrieb der Anwendung gewährleistet.

Das ETH Data Archive wurde innerhalb der ETH Zürich vor allem unter dem Gesichtspunkt der Sicherung der dauerhaften Verfügbarkeit angeboten. Für Forschende, die ihre Daten publizieren wollten, war insbesondere die zitierbare Veröffentlichung mit Digital Object Identifier (DOI)⁷ ein wichtiges Argument.

Zur Nutzung des ETH Data Archive wurden Workflows mit unterschiedlicher Charakteristik und verschiedenen Automatisierungsgraden angeboten. Für eher kleinteilige, dateibasierte Daten des sogenannten ‚Long Tail‘ stand den Kundinnen und Kunden ein manueller, browserbasierter Upload zur Verfügung. Er wurde primär für die Ablage von Datenpaketen verwendet, die ergänzend zu einer Publikation online und zitierbar zugänglich gemacht werden sollten (‚Supplementary Material‘). In diesem Workflow konnte pro Metadatensatz nur eine Datei oder ein Container im ZIP-Format abgeliefert werden.

Docuteam packer für ‚Small Data‘

Als ergänzendes Werkzeug wurde daher der docuteam packer⁸ gewählt. Dabei handelt es sich um einen Viewer und Editor, der ursprünglich von der Firma docuteam für den Archivsektor entwickelt und als Open Source Software publiziert wurde. Er wurde einerseits gewählt, um die Workflows des Hochschularchivs der ETH Zürich zu unterstützen, und wurde andererseits durch erhebliche Weiterentwicklungen zusätzlich auf Bedürfnisse von Forschungsgruppen ausgerichtet. Die Stärke des docuteam packer liegt in der Möglichkeit, lokal eine Ordnerstruktur aufzubauen und auf jeder Ebene mit Metadaten zu versehen. Dabei werden bereits DOI reserviert, die später im ETH Data Archive registriert werden. Zudem liefert er Informationen zu Dateiformaten und erleichtert die Bearbeitung durch eine Previewfunktion.

Während der Bearbeitung im docuteam packer bleiben die Daten in der von den Forschenden gewählten Umgebung gespeichert und können z.B. über die Dauer einer Doktorarbeit hinweg gesammelt und kontinuierlich dokumentiert werden. Erst wenn der oder die Forschende die Sammlung in Teilen oder insgesamt als abgeschlossen betrachtet, kann sie oder er das betreffende Paket per Knopfdruck an das ETH Data Archive abschicken. Aufgrund der hinterlegten Konfiguration wird das Paket dann zunächst durch die zentral gehostete Anwen-

⁷ <http://www.library.ethz.ch/DOI-Desk>.

⁸ <https://www.docuteam.ch/angebot/archivinformatik/software/>.

derung docuteam feeder weitgehend automatisch aufbereitet und dem ETH Data Archive übergeben. Neben der Formatanalyse und -validierung nimmt das ETH Data Archive die DOI-Registrierung vor und publiziert die abgelieferten Metadaten im Wissensportal⁹ der ETH-Bibliothek (Anwendung: Ex Libris Primo). Für die Bedürfnisse des Hochschularchivs weicht der Weg ab: Die Metadaten werden in das Archivinformationssystem (Anwendung: CMI Star) publiziert und gelangen von dort ins Wissensportal. Die entsprechenden Workflows werden jedoch derzeit erst in Betrieb genommen, unter anderem für die Erschliessung der WARC (Web Archive)-Dateien aus der Archivierung von Webseiten der ETH Zürich.

Nach heutigem Stand gehen wir davon aus, dass der docuteam packer vom Hochschularchiv und verwandten Kunden intensiv genutzt werden und für die Bearbeitung von Forschungsdaten eher ein Nischenprodukt bleiben wird. Sein Nutzen wird von Kunden mit entsprechenden Bedürfnissen sehr schnell erkannt. Die Erfahrung zeigt aber, dass Forschende die – transparent kommunizierte – Grenze einer annehmbaren Performance schneller erreichen, als sie selbst erwarten. Diese liegt derzeit bei etwa 2 Gigabyte pro Paket, wobei die Anzahl der Pakete lediglich durch den verfügbaren Speicherplatz beschränkt ist.

Automatisierung

Als dritte Variante für die Übergabe von Daten an das ETH Data Archive wird ein automatischer Transfer von Server zu Server mit Hilfe sogenannter ‚Submission Applications‘ unterstützt. Diese können eingesetzt werden, wenn aus einer Datenquelle regelmässig ein wohldefinierter Output von Metadaten und Dateien exportiert werden kann, der archiviert werden soll. Damit ist bereits gesagt, warum diese Art von Workflow bisher ausschliesslich für solche Quellanwendungen zum Einsatz kommt, die von der ETH-Bibliothek gepflegt werden und somit nach gut bekannten Standards arbeiten und grosse Mengen gleichartiger Dateien ausgeben. Die Anbindung an bestehende Plattformen für das Forschungsdatenmanagement wird weiterverfolgt werden, sobald entsprechende Anwendungsfälle identifiziert werden konnten.

Die Workflows für Forschungsdaten im ETH Data Archive wurden zunächst ohne Verbindung zur ETH E-Collection, dem institutionellen Repositorium der ETH Zürich, aufgebaut. Diese offensichtlich nicht ideale Trennung war unter anderem dadurch begründet, dass eine grundlegende Erneuerung der E-Collection bereits in Vorbereitung war und keine umfangreicheren Eingriffe in die noch

⁹ <http://www.library.ethz.ch>.

laufende Anwendung mehr vorgenommen werden sollten. Bei der Ausschreibung und Implementierung einer Nachfolgelösung wurden die für Forschungsdaten benötigten Workflows dann von vornherein mitkonzipiert. Die aktuelle Lösung wird als Teil der Sicht auf die Gesamtsituation 2017 erläutert.

Datenmanagementpläne

Mit dem wachsenden Verständnis für das Thema Forschungsdatenmanagement und seine Herausforderungen wurden in den 2010er Jahren zunehmend die Organisationen der Forschungsförderung aktiv. Im Sinne einer möglichst effizienten Nutzung der von ihnen eingesetzten Fördermittel begannen sie, die Empfängerinnen und Empfänger ihrer Förderung zu verschiedenen Aktivitäten zu ermuntern und zu verpflichten. Um die Qualität des Umgangs mit Forschungsdaten als wesentlichem Teil der guten wissenschaftlichen Praxis zu verbessern, werden Datenmanagementpläne (DMP) als Teil von Projektanträgen verlangt, zuletzt etwa vom Schweizerischen Nationalfonds (SNF) für Projektanträge ab Oktober 2017.¹⁰ Neben der Behebung vermuteter Mängel im Forschungsdatenmanagement und eines möglichst effizienten Mitteleinsatzes, der die mehrfache Erzeugung gleicher Daten vermeidet, rückt zunehmend das Interesse in den Mittelpunkt, Daten explizit zu publizieren und als Open Data zur Verfügung zu stellen. Die Ankündigung des SNF verunsichert aktuell einige Forschende und die Fachstelle Digitaler Datenerhalt versucht in Rücksprache mit dem SNF, Forschenden der ETH Zürich die Vorgaben im Detail zu erläutern und andererseits auch dem SNF deren Rückmeldungen zukommen zu lassen.

Bestandsaufnahme 2017: Wo stehen wir?

Das Jahr 2017 brachte eine wesentliche Neuausrichtung der Dienstleistungen der ETH-Bibliothek zum Forschungsdatenmanagement. Mitte des Jahres nahm die Research Collection¹¹ ihren Betrieb auf. Sie ist die neue Publikationsplattform für alle Arten von Forschungsooutput von Angehörigen der ETH Zürich (Anwendung: DSpace¹²). Sie löst damit die vorherigen Anwendungen E-Collection (institutionelles Repositorium) und E-Citations (Hochschulbibliographie mit Anbindung an

¹⁰ <http://www.snf.ch/de/fokusForschung/newsroom/Seiten/news-170511-open-research-data-snf-guidelines-fuer-datenmanagementplaene.aspx>.

¹¹ <https://www.research-collection.ethz.ch/>.

¹² <http://www.dspace.org/>.

das Reporting) ab und ergänzt diese um die Möglichkeit, Forschungsdaten zusammen mit Publikationen oder auch eigenständig zu publizieren. Dieser Workflow für ‚Supplementary Material‘ war bisher im ETH Data Archive angesiedelt, konnte dort aber nicht überzeugend und mit der nötigen Flexibilität implementiert werden. Hauptgrund dafür sind die Restriktionen, die mit dem Anspruch der Langzeitarchivierung einhergehen. Während dieser Workflow unter Nutzung einer Weboberfläche im ETH Data Archive nicht mehr weitergeführt wird, wird künftig der gesamte Inhalt der Research Collection über eine Schnittstelle in das ETH Data Archive importiert. Dies war seit 2012 bereits für die E-Collection der Fall. Das ETH Data Archive fungiert hier als ‚Dark Archive‘, auf das im Normalfall nur die Mitarbeitenden zugreifen können, die die Anwendung betreiben.

Für die Kundinnen und Kunden steht somit nun eine einzige Plattform zur Verfügung, um sowohl Publikationen und Referenzen als auch Forschungsdaten publizieren und ablegen zu können. Dies ist insbesondere wichtig im Hinblick auf die Forderung des SNF, Daten aus den von ihm geförderten Projekten gemäss den FAIR Prinzipien¹³ zugänglich zu machen. Auch weiterhin können Forschende wählen, ob sie in einem fachspezifischen Repository Daten hinterlegen möchten, das den Kriterien genügt. Es ist zudem klar, dass eine Online-Plattform wie die Research Collection primär für den Upload von überschaubaren Datenmengen via Browser gedacht ist. Je nach Browser sind dies derzeit nur wenige Gigabyte, so dass weiterhin nur der ‚Long Tail‘ von Forschungsdaten bedient wird. Für grössere Datenmengen können Metadaten als Nachweis erfasst werden, sofern eine stabile Referenz besteht. Falls solche umfangreicheren Daten noch gar nicht definitiv abgelegt sein sollten, kann weiterhin der Transfer ins ETH Data Archive von Server zu Server über eine Submission Application geprüft werden.

Aufgabenteilung

Parallel beschäftigen sich die ETH-Bibliothek und ihre Partner mit weiteren Fragen rund um die Unterstützung des Forschungsdatenmanagements. Während die Diskussion in der Anfangszeit vor allem mit den für die Speicherinfrastruktur verantwortlichen Kollegen geführt wurde, sind längst auch andere Gruppen involviert.¹⁴ Konkret sind dies die Scientific IT Services (ID SIS) der Informatikdienste der ETH Zürich, deren Dienstleistungen vor allem die Erzeugung und

¹³ <https://www.force11.org/group/fairgroup/fairprinciples>.

¹⁴ Ana Sesartić, Matthias Töwe: Research Data Services at ETH-Bibliothek. In: IFLA Journal 42/4 (2016), S. 284–291 (<https://doi.org/10.1177/0340035216674971>).

Verarbeitung von Forschungsdaten während ihrer aktiven Nutzung im Forschungsprozess unterstützen. Dazu gehören Entwicklung und Pflege von spezialisierter Software ebenso wie die Bereitstellung von Ressourcen für das Hochleistungsrechnen (High Performance Computing, HPC). Aus Sicht von Kundinnen und Kunden ist natürlich anzustreben, dass sie möglichst wenige verschiedene Stellen ansprechen müssen, um die benötigten Dienstleistungen nutzen zu können. Eine Herausforderung besteht somit zunächst darin, innerhalb der jeweils gewachsenen Strukturen die relevanten Informationen zusammenzufassen und an einem Ort bereitzustellen. Einige Schulungsangebote werden bereits gemeinsam von Kolleginnen und Kollegen der ETH-Bibliothek und der Scientific IT Services durchgeführt. Weitere Aktivitäten unter anderem zu Datenmanagementplänen werden folgen. Aus technischer Sicht wird es darum gehen, aus den Tools für das Management aktiver Forschungsdaten möglichst unkompliziert Daten für eine Veröffentlichung und für die dauerhafte Erhaltung auswählen und exportieren zu können.

Gemeinsames Ziel der Bemühungen aus technischer und organisatorischer Sicht ist es, die Bereitstellung von Forschungsdaten als Open Data durch die möglichst umfassende Umsetzung der FAIR-Prinzipien (Findable – Accessible – Interoperable – Re-usable) zu fördern. Diese Prinzipien können nicht erst durch die Ablage in einem geeigneten Repositorium erfüllt werden, sondern ihre verschiedenen Aspekte müssen frühzeitig durch die nachvollziehbare Dokumentation des Entstehungs- und Bearbeitungsprozesses vorbereitet werden.

Eine wichtige Rolle jenseits der technischen Infrastrukturen spielt der Stab Forschung der ETH Zürich, bei dem eine Vielzahl von Funktionen angesiedelt sind, die Forschende zum Beispiel bei der Antragstellung für Forschungsprojekte, bei der Abklärung forschungsethischer Fragen sowie bei weiteren Fragen mit Relevanz für die Planung und Durchführung von Forschungsprojekten unterstützen. Dem Stab Forschung kommt wie seinen Pendants in anderen Hochschulen eine Schlüsselrolle dabei zu, Richtlinien für den Umgang mit Forschungsdaten zu definieren und ihre Einhaltung voranzutreiben und durchzusetzen. Eine Herausforderung besteht sicher darin, die Interessen der Hochschule, der Forschenden, der Forschungsförderer, der weiteren Öffentlichkeit sowie weiterer Interessengruppen angemessen zu berücksichtigen. Hier bleibt trotz der bestehenden allgemeinen Verpflichtungen aus den Richtlinien für die gute wissenschaftliche Praxis noch einiges zu tun.

Open Data – FAIR Data

Auf einer anderen Ebene geht es auch darum, neue Anreize zu schaffen, die das Forschungsdatenmanagement positiv beeinflussen können, und bestehende Mechanismen zu hinterfragen, die die Anerkennung von Forschungsleistungen einseitig über Anzahl und angeblichen Impact von formalen Publikationen definieren. Ein wichtiger Schritt in diese Richtung sind die Vorschläge gemäss der DORA-Deklaration¹⁵, die in ihrer Zielrichtung und ihren möglichen Auswirkungen selbstverständlich weit über das Thema Datenmanagement hinausgehen.

Es muss allerdings auch angemerkt werden, dass die teilweise weitgehenden Forderungen nach dem offenen Zugang zu Forschungsdaten insbesondere bei den Forschenden selbst nicht unumstritten sind. Dies hat verschiedene Gründe. Einerseits erscheint es als technisch und wirtschaftlich nicht vertretbar, die in manchen Fächern anfallenden grossen Datenmengen pauschal für Dritte vorzuhalten, womöglich sogar zur Selbstbedienung. Andererseits zweifeln nicht wenige Forschende daran, dass es wirklich generell möglich ist, Forschungsdaten so umfassend zu dokumentieren, dass eine wissenschaftlich seriöse Nachnutzung ohne Rücksprache mit den Datenproduzenten möglich ist. Je näher man diesem Ideal kommt, desto höher ist der Aufwand, der investiert werden muss und für die eigene Forschungsarbeit fehlt. Zudem fehlt es bisher schlicht an Erfahrung, welche Daten man allgemein zugänglich machen kann und welche beispielsweise eher auf Anfrage aufgrund vorhandener Metadaten.

Ein Grundproblem ist sicher, dass sowohl Geldgeber als auch Institutionen Vorgaben machen müssen, die für alle Disziplinen gelten, während die genaue Umsetzung sehr disziplinspezifisch erfolgen muss. Damit liegt eine grosse Verantwortung bei den jeweiligen Fachcommunities, die ihre eigenen Erwartungen aneinander definieren müssten, aber diese nicht einfache Aufgabe in der Masse noch nicht aufgegriffen haben.

Wie geht es weiter?

Die ETH Zürich wird ihre Infrastrukturen für das Forschungsdatenmanagement weiterentwickeln, sowohl unter Gesichtspunkten der guten wissenschaftlichen Praxis als auch im Hinblick auf die Bereitstellung ausgewählter Daten gemäss den FAIR-Prinzipien. Dabei müssen und werden die verschiedenen Beteiligten innerhalb der Hochschule in Zukunft mit grosser Wahrscheinlichkeit noch enger

15 DORA: San Francisco Declaration on Research Assessment (<http://www.ascb.org/dora/>).

zusammenarbeiten als bisher: Die anstehenden Aufgaben machen an gewachsenen Grenzen von Organisationseinheiten keinen Halt. Die Verteilung von Aufgaben auf mehrere Schultern erfordert naturgemäss einigen Koordinationsaufwand, sie ist aber auch eine grosse Chance, eine tatsächlich umfassende Abdeckung der Themen zu erreichen.

Nächster greifbarer Schritt könnte die im Grunde überfällige Formulierung einer Datenpolicy bzw. eines vergleichbaren Dokuments sein, um die bestehenden pauschalen Richtlinien bis zu einem gewissen Grad zu präzisieren.

Die Zusammenarbeit geht heute bereits weit über die Grenzen der ETH Zürich hinaus. Im Projekt Research Data Life-Cycle Management (DLCM)¹⁶ im Programm P-5 „Wissenschaftliche Information“¹⁷ von swissuniversities arbeiten acht Institutionen auf verschiedenen Ebenen an der Unterstützung des Forschungsdatenmanagements.¹⁸ Die Bandbreite der Teilprojekte reicht von gemeinsamen Richtlinien, Policies und Schulungsinhalten über Plattformen für das Management von Forschungsdaten im aktiven Forschungsprozess bis hin zu Lösungen für die Langzeitarchivierung, die idealerweise als Services angeboten werden sollten. Die Herausforderungen sind beträchtlich und es wird sich zeigen, welche Aufgaben sich tatsächlich für eine Zentralisierung oder kooperative Bewältigung eignen. Eine nicht zu unterschätzende Schwierigkeit stellt die Formulierung tragfähiger Geschäftsmodelle für die unterschiedlichen Dienste dar, insbesondere da einige andernorts übliche Ansätze für solche Modelle in und zwischen den Hochschulen bisher zumindest ungebräuchlich waren und dementsprechend schlecht mit bisherigen Prozessen und Strukturen vereinbar sind.

16 <https://www.dlcm.ch/>.

17 <https://www.swissuniversities.ch/de/organisation/projekte-und-programme/p-5/>.

18 Eliane Blumer, Pierre-Yves Burgi: Data Life-Cycle Management Project: SUC P2 2015-2018. In: RESSI 16 (2015), Art. 110 (http://www.ressi.ch/num16/article_110).