

Der Umgang mit heterogenen (Forschungs-)daten an einer wissenschaftlichen Bibliothek – Use Cases und Erfahrungen aus technischer und nichttechnischer Sicht an der Universität Wien

Susanne Blumesberger, Raman Ganguly

ZID Universität Wien

Zusammenfassung

Forschungsdatenmanagement an wissenschaftlichen Bibliotheken setzt sich aus mehreren Aspekten zusammen. Neben der reinen administrativen Arbeit, der Bereitstellung einer Möglichkeit, die Daten kurz-, -mittel oder langfristig verfügbar zu machen, sind auch technische und juristische Expertisen notwendig. Der Beitrag schildert anhand des Beispiels der Universität Wien worauf beim Umgang mit Forschungsdaten geachtet werden sollte.

Abstract:

Research data management at scientific libraries consists of several aspects. In addition to pure administrative work, the provision of a way to make the data short, medium or long-term available, also technical and legal expertise is necessary. The article uses the example of the University of Vienna to describe what should be considered when dealing with research data.

Das Repositorium als Ausgangspunkt für Forschungsdatenmanagement

Gemeinsam mit dem Zentralen Informatikdienst der Universität Wien startete die Universitätsbibliothek Wien 2007 mit dem Aufbau eines universitätsweiten Repositories, das auf der Open Source Software Fedora [1] basiert. Unter Einbeziehung mehrerer Universitätsinstitute wurden Anforderungen aus unterschiedlichen Wissenschaftsbereichen erhoben und in die Planung und Ausarbeitung miteinbezogen. Der Fokus für das Repositorium war schon bereits bei der Konzeption nicht nur ausschließlich auf die Forschung gerichtet sondern auch auf die Lehre. Als Metadatenschemata wurden das LOM-Schema (Learning Object Metadata) und Dublin Core gewählt. Bei der Entwicklung der Software arbeiten die drei Abteilungen: Universitätsbibliothek, Zentraler Informatikdienst und Center for Teaching und Learning eng zusammen.

Im April 2008 ging [PHAIDRA](#) (ein Akronym für Permanent Hosting, Archiving and Indexing of Digital Resources and Assets) online. Von Beginn an stand das Repositorium allen MitarbeiterInnen und Studierenden der Universität Wien offen, zunächst mit begrenztem Speicherplatz, später ohne Limitierung. Zusätzlich können Zugänge für externe Forschende vergeben werden, was ein kollaboratives Arbeiten mit dem System möglich macht. Ergänzt wurde PHAIDRA schon bald durch [u:scholar](#), den Institutionellen Repositorium der Universität Wien, das mit einer eigenen Upload- und Suchmöglichkeit für wissenschaftliche Publikationen ausgestattet sind. Archiviert werden auch diese Objekte in PHAIDRA. Auf die rechtliche Sicherheit wurde von Beginn an geachtet, so stehen unter anderem den Usern CC-Lizenzen zur Verfügung. Objekte ohne Kennzeichnung des Urheberrechtes sind nicht zulässig. Allen NutzerInnen von PHAIDRA wurden bei der ersten Benutzung die [Nutzungsbedingungen](#) zur Kenntnis gebracht, die vorsehen, dass nur Objekte archiviert werden dürfen, die aus urheberrechtlicher Sicht auch in das Repositorium abgelegt werden dürfen und gegen die auch sonst keine juristischen oder ethischen Gründe vorliegen.

In PHAIDRA können alle Arten von Objekten archiviert werden, unter anderem Texte, Bilder, Video- und Audiofiles sowie Daten. Derzeit sind neben Artikeln, Bilder, die aus der Forschung entstanden sind und/oder der Lehre dienen, Karten, Bücher, Audio- oder Videomittschnitte von Vorlesungen, Präsentationen, Grafiken, Statistiken, alle Arten von Sammlungsobjekten und vieles mehr archiviert. Die Owner, jene Personen, die Objekte hochladen, können die Beschreibungen, Metadaten, immer wieder ergänzen und verändern. Die Objekte selbst können, da dies der Policy der Langzeitverfügbarkeit widersprechen würde, nicht mehr gelöscht werden. Allerdings ist eine Versionierung möglich, die neue Version erhält einen neuen Permalink und wenn gewünscht auch eine eigene Beschreibung. Wichtig für die Forschenden ist auch, dass der Zugang zu den Objekten zunächst zwar - entsprechend der Open Access Policy der Universität Wien - weltweit geöffnet ist, durch den Owner jedoch sofort auf die gesamte Universität Wien, auf Institute, Departments, Personengruppen und Einzelpersonen gesperrt werden kann. Diese Sperre kann mit einem Embargo versehen werden und jederzeit verändert werden. Damit ist es zum Beispiel möglich innerhalb eines Forschungsprojekts zunächst unter Ausschluss der Öffentlichkeit zu arbeiten und nach Projektende Objekte automatisch freizugeben. Diese Funktion

macht es auch möglich, Artikel in PHAIDRA zu archivieren, die vom Verlag aus noch unter einer Embargozeit stehen, also erst nach einer gewissen Zeitspanne verbreitet werden dürfen. Zu den fast seit Beginn bestehenden Services zählt auch die Möglichkeit mittels eines PHAIDRA-Importers Collections und eigene Bücher zu erstellen und im Bookviewer anzeigen zu lassen. Die Funktion der Collections ermöglicht die gemeinsame Darstellung von unterschiedlichen Objekten. Collections erhalten einen eigenen Permalink und eine eigene Beschreibung. Jede Collections kann wiederum Collections enthalten, was beispielsweise die Darstellung von Zeitschriften nach Jahrgängen bis auf Artekelebene ermöglicht. Das Besondere daran ist, dass Collections auch von Objekten gebildet werden kann, von denen man nicht der Owner ist. Damit wird es beispielsweise möglich per Link einen ganzen Satz an Objekten für eine Lehrveranstaltung an Studierende zu schicken.

Genutzt wird PHAIDRA derzeit von nahezu allen Disziplinen, vor allem aber von den Geisteswissenschaften [2], was unter anderem zeigt, dass diese Disziplinen bisher keine geeigneten Tools hatten.

Neue Anforderungen der User

Im Laufe der Zeit veränderten sich die Anforderungen der NutzerInnen. Stand am Anfang vor allem die sichere Langzeitverfügbarkeit im Mittelpunkt, kamen neue Bedürfnisse hinzu, wie etwa die Frage nach kurz-, bzw. mittelfristigen Speichermöglichkeiten. Neben der unterschiedlichen Dauer der Archivierung wird vermehrt der Bedarf an Archivierung von Datenbanken und Software erkennbar.

Aufgrund der neuen Herausforderungen wurde die Architektur von einem Repositorium hin zu einem Ökosystem aus Repositorien und Daten-Services geändert. So wurde zum Beispiel die Landschaft durch ein Git basierendes Repositorium [3] ergänzt um Software Code und dynamische Daten abzulegen. An diesen Anforderungen lässt sich auch ein verändertes Forschungsverhalten ablesen.

Als Services für die Daten werden automatisch handle IDs [4] für jedes Objekt mitgeliefert, zusätzlich werden DOI's für die Objekte auf Nachfrage über ein eingerichtetes DOI-Service vor oder nach der Archivierung vergeben.

Neben den Services für die Zitierung von Daten, werden die Visualisierung der Forschungsdaten und die Darstellung der Ergebnisse auf Projekt bezogen Webseiten immer wichtiger. So können Daten von Phaidra über Schnittstellen an in andere Webauftritte integriert werden. Dabei orientiert sich die technischen Umsetzung an den Empfehlung von der Vereinigung Confederation of Open Access Repositories (COAR), die im Rahmen der Interest Group Next Generation of Repositories [5]. Über diese Möglichkeit haben wir bereits Forschungsprojekt in ihrer Darstellung in der Öffentlichkeit unterstützt.

Auch unsere Projektpartner nutzen diese Möglichkeit um Applikationen für den Ingest und den Re-use an PHAIDRA anzubinden. Grundgesetzlich ist unser Designparadigma als ein System in der Mitte, in dem auch Maschinen, nicht nur Menschen die Daten ablegen und wiederverwenden können. Es ist ein gut organisierte Lagerhalle von Daten, nicht deren Produktionsort und auch nicht der Ort an dem sie visualisiert werden.

Dennoch bieten wird im Rahmen unserer Services auch Darstellung für die spezielle Datentypen an. So können Videofiles von PHAIDRA aus gestreamt werden und für Bilder habe wir einen Image Server. Mithilfe dieses Servers können hochauflösende Bilder über das Web zoombar dargestellt werden und die Bilder können über den IIIF Standard (International Image Interoperability Framework) [6] angesprochen werden.

Parallel zu den Forderungen nach einer sicheren Archivierung der Daten wuchs die Nachfrage von Beratung vor allem im technischen und im juristischen Bereich. Das hängt sicher auch damit zusammen, dass einige Fördergeber bereits Datenmanagementpläne verlangen und einen bewussten und nachhaltigen Umgang mit Forschungsdaten einfordern. Der [FWF](#) (Fonds zur Förderung der wissenschaftlichen Forschung) verlangt zum Beispiel bei aktuellen Einreichungen von Forschungsprojekten in seiner [DMP-Vorlage](#) Angaben über Datennutzungs- und Speicherungsstrategien sowie Hinweise auf den juristischen und ethischen Umgang mit Daten.

Datenmanagement anhand von Datenmanagementplänen

Verpflichtende Datenmanagementpläne haben dazu geführt, dass Forschende zunehmend Unterstützung bei der Antragstellung benötigen. Diese Unterstützung kann nur bis zu einem gewissen Grad durch Schulungen und Präsentationen abgedeckt werden, denn Forschungsdaten sind sehr unterschiedlich, betreffend die Struktur, die Menge aber auch was juristische und ethische Fragen betrifft. Dies wiederum verlangt den Einsatz unterschiedlicher Tools und Strategien. Während auf der einen Seite Open Science immer stärker eingefordert wird, vor allem von den Fördergebern, zeigt der langjährige Kontakt mit Forscherinnen und Forschern aus verschiedenen Fachbereichen, dass in einigen Bereichen sehr sensibel mit Daten umgegangen werden muss. Dazu zählen nicht nur Ergebnisse aus medizinischen Projekten sondern beispielsweise auch aus sozialwissenschaftlichen und geisteswissenschaftlichen Zusammenhängen. So kann beispielsweise ein vor 20 Jahren geführtes Interview im Rahmen der Kultur- und Sozialanthropologie mit einer kleinen Personengruppe, diese heute aufgrund veränderter politischer Gegebenheiten gefährden. Aber auch unklare rechtliche Besitzverhältnisse von Daten beeinträchtigen die Vision einer offenen Wissenschaft. Zu diesen extern bedingten Einschränkungen kommen auch Unsicherheiten und Widerstände bei den Forscherinnen und Forschern hinzu, die sich vor einer allzu freien Preisgabe ihrer Daten scheuen. Die Gründe dafür sind vielfältig. Bei hochkompetitiven Fachbereichen ist die Angst vor Ideenklau vorherrschend, in anderen Bereichen überwiegt wiederum die Sorge, dass Daten missbräuchlich verwendet werden könnten. Man denke dabei nur an Forschungen über den Nationalsozialismus.

Neue Herausforderungen

Die Herausforderungen für TechnikerInnen und BibliothekarInnen besteht nun darin, die Forschenden in einer Art und Weise zu unterstützen, dass ihnen hinsichtlich Verfügbarmachung ihrer Daten möglichst viele Optionen offenstehen. Das bedeutet, sie benötigen auf der technischen Seite einerseits offene Repositorien, in die sie Daten hochladen und weltweit mit einer DOI versehen zugänglich machen können, andererseits aber auch geschlossene Systeme, die es ihnen erlauben, ihre Ergebnisse sicher zu verwahren und vor unberechtigten Zugriffen zu schützen. Das betrifft zum Teil nicht nur die Daten sondern auch die Metadaten. Für manche Daten ist auch die Lösbarkeit unbedingt erforderlich, vor allem im medizinischen Bereich. Auf der

nichttechnischen Seite wiederum sind Strukturen notwendig, die entlang des gesamten Forschungsprozesses Beratungen zu juristischen oder ethischen Fragestellungen anbieten, bzw. zum Beispiel hinsichtlich Metadatenvergabe unterstützend wirken.

Eine weitere Herausforderung ist, wie oben bereits erwähnt wurde, dass Forscherinnen und Forscher nicht nur Daten in Form von Dateien archivieren wollen, sondern auch Datenbanken und Software. Bei diesen beiden Datenklassen versagt die herkömmliche Archivierung in einem Repositorium, da der Kontext und die Umgebung für die Lauffähigkeit ein wesentlicher Bestandteil der Archivierung sein muss. Ohne diese beiden Teile kann der Zustand in der Zeit zwar eingefroren werden, es kann aber nicht mehr von einer langfristigen Zurverfügungstellung von Daten gesprochen werden.

Datenmanagement als Teamwork

An der Universität Wien blicken wir nun auf eine mehr als zehnjährige Erfahrung im Bereich Forschungsdatenmanagement zurück. Der Zentrale Informatikdienst der Universität Wien arbeitet dabei sehr eng mit der Universitätsbibliothek zusammen, betreibt nicht nur gemeinsam das universitätsweites Repositorium PHAIDRA, sondern führt vor allem Beratungsgespräche immer im Team durch um auf die unterschiedlichen Anforderungen rasch eingehen zu können. Durch die 19 Partneruniversitäten von PHAIDRA hat sich ein großes [Netzwerk](#) gebildet, das sich in technischen und auch in nichttechnischen Belangen gegenseitig stützt. Das über 90 Personen umfassende [RepositorienmanagerInnennetzwerk](#) (RepManNet) tauscht zusätzlich systemübergreifend Erfahrungen aus arbeitet in diversen Arbeitsgruppen an der Konzeption von zukünftigen Repositorien und an Guidelines. Den Forschenden an der Universität Wien steht zusätzlich das Netzwerk "[code4research](#)" zur Verfügung, das dem Erfahrungsaustausch der WissenschaftlerInnen bzgl. der langfristigen Verfügbarkeit von komplexeren Daten wie Datenbanken und Software dient.

Zusammenfassend lässt sich sagen, dass Forschungsdatenmanagement nur im Zusammenspiel durch mehrere Personen mit unterschiedlichen Expertisen funktioniert, dass eine enge und von Vertrauen geprägte Zusammenarbeit mit den Forschenden erfolgen muss und dass die beteiligten Personen sich ständig auf neue

Herausforderungen einlassen müssen. Dabei muss auch mit anderen Stellen der Universität, zum Beispiel mit dem Forschungsservice, aber auch mit den Fördergebern gut zusammengearbeitet werden. Diese Zusammenarbeit führt nicht nur zu einem Ausbau der Kompetenzen auf allen Seiten, sondern auch zu konkreten Verbesserungen der Systeme. Für PHAIDRA hat dies beispielsweise im letzten Jahr bedeutet einen neuen Objekttyp, einen Container mit komplexen Features, zu entwickeln.

Fußnoten

[1] Die Software hieß zu den damaligen Zeitpunkt noch Fedora Commons. Die Umbenennung erfolgte im Zuge der Zusammenführung der Entwicklungen der beiden großen Open Source Repositorien Fedora und DSpace. Für beide ist nun die Organisation DuraSpace verantwortlich. Weiterführende Informationen sind online zu finden unter: <https://duraspace.org/>

[2] Bei der letzten statistischen Erhebung von den Objekte in PHAIDRA im Jänner 2019 machen Objekte die der Digital Humanities zugeordnet werden können 65% aus.

[3] Zur Zeit verwendet die Universität Wien GitHub Enterprise Git Repository.

[4] Handle ist die technische Basis von DOI und ist Open Source. Weiter Information sind online zu finden unter: <https://www.handle.net>

[5] „Our vision is to position repositories as the foundation for a distributed, globally networked infrastructure for scholarly communication, on top of which layers of value added services will be deployed, thereby transforming the system, making it more research-centric, open to and supportive of innovation, while also collectively managed by the scholarly community. However, in order to leverage the value of the repository network, we need to equip it with a wider array of roles and functionalities, which can be enabled through new levels of web-centric interoperability.“ ist das Ziel der Arbeitsgruppen Next Generation Repository von COAR, die auf internationalen Experten der Bibliothek und der IT besteht. Online zu finden unter: <https://www.coar-repositories.org/activities/advocacy-leadership/working-group-next-generation-repositories/>

[6] Weitere Informationen zu diesem Framework könne online unter <https://iif.io/> abgerufen werden.

Literatur

Blumesberger, Susanne: Neue Anforderungen – viele offene Fragen. Zu den vielfältigen Rollen von Repositorien am Beispiel der UB Wien. In: [Mitteilungen der VÖB](#)

Eberhard, Igor; Kraus, Wolfgang: Der Elefant im Raum. Ethnographisches Forschungsdatenmanagement als Herausforderung für Repositorien. In: [Mitteilungen der VÖB](#)

Ganguly, Raman: Nachhaltige Softwareentwicklung: code4research. In: [Mitteilungen der VÖB](#)

Ganguly, Raman, Paolo Budroni, and Barbara Sánchez Solís. "Living Digital Ecosystems for Data Preservation: An Austrian Use Case Towards the European Open Science Cloud." In Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices, 203-210. ELPUB. Limassol, Cyprus, 2017; <https://elpub.architexturez.net/doc/10-3233/978-1-61499-769-6-203>

Hagmann, Dominik: Überlegungen zur Nutzung von PHAIDRA als Repository für digitale archäologische Daten. In: [Mitteilungen der VÖB](#)

Kulturpool: Österreichs Portal zu Kunst, Kultur und Bildung (30. April 2018): [Phaidra, die Strahlende](#). Interview mit Susanne Blumesberger.

Torggler, Andrea; Andrae, Magdalena: Aus dem Leben einer/s Repman – ein Bericht aus dem österreichischen „Netzwerk für RepositorienmanagerInnen“. In: [Mitteilungen der VÖB](#)