

Langzeitarchivierung von Forschungsdaten in den Geisteswissenschaften.

Bedürfnisse von Forschenden und Möglichkeiten der Universitätsbibliothek Bern

Ursula Loosli

Neue Forderungen, neue Bedürfnisse?

Forschende werden in den letzten Jahren durch Auflagen von forschungsfördernden Institutionen mit einschneidenden Veränderungen konfrontiert. Grundlage für diese Auflagen ist die Open Access¹ Bewegung. Während eine Forschungsdaten-Policy an der Universität Bern noch aussteht, stellen andere Institutionen bereits konkrete Forderungen: Nicht nur sollen bis 2020 alle vom Schweizerischen Nationalfonds (SNF) geförderten Publikationen Open Access zur Verfügung stehen, seit Oktober 2017 betrifft dies auch Forschungsdaten, sofern ihrer Veröffentlichung keine rechtlichen, ethischen oder urheberrechtlichen Gründe widersprechen.² Beim Einreichen des Projektantrags wird verlangt, einen Data Management Plan (DMP) auszufüllen, der unter anderem festhält, wann und wo welche im Projekt entstehenden Daten publiziert und archiviert werden sollen. Forschende, die sich bisher nicht mit der Thematik befasst haben, stehen somit vor einer gewaltigen Aufgabe, die viele Fragen und Unsicherheiten mit sich bringt.

Ein Bereich davon betrifft die digitale Langzeitarchivierung (LZA).³ Als problematisch zu betrachten ist, dass die bestehenden Angebote für die LZA « [...] die sich abzeichnende breite Anforderungspalette geisteswissenschaftlicher digitaler Forschung noch nicht vollständig»⁴ erfüllen.

Die Aktualität des Themas LZA zeigt sich beispielsweise an Veranstaltungen wie der Tagung «Editions- und Forschungsplattformen zum 18. Jahrhundert» des Historischen Instituts der Universität Bern vom 8. Mai 2019. Nachhaltigkeit und LZA tauchten als Leitthemen immer wieder auf und Teilnehmende der Konferenz bezeichneten Universitätsbibliotheken in diesen Bereichen als potenzielle Partner. Dies ist berechtigt, da die Gewährleistung eines langfristigen Erhalts von Informationen und

1 Die SAGW definiert Open Access als „sofortigen, permanenten, freien, kostenlosen und elektronischen Zugang zu wissenschaftlichen Publikationen“, vgl. SAGW (o. J. a).

2 Vgl. SNF (o. J. a).

3 LZA steht im Folgenden für Langzeitarchivierung und wird synonym zum Begriff ‚digitale Langzeitarchivierung‘ verwendet.

4 DHd AG Datenzentren (2017), S. 5.

auch von deren Benutzbarkeit eine traditionelle Kernaufgabe von Universitätsbibliotheken darstellt und Forschende eine der Hauptzielgruppen der Universitätsbibliothek Bern (UB Bern) sind. In diesem Sinne will die UB Bern Forschende bei neuen Aufgaben begleiten und dazu das Dienstleistungsangebot erweitern, unter anderem im Bereich der LZA.

Aber welche Unterstützung benötigen die Forschenden? Was soll und kann die UB Bern in Bezug auf Fragen der LZA von Forschungsdaten leisten?

Fragestellung und Methodik der Masterarbeit

Die Bedürfnisse der eigenen Forschenden an der UB Bern waren bisher unbekannt. Deshalb hat die Masterarbeit, die vorliegendem Artikel vorangeht, einen ersten Vorstoss im Bereich der Geisteswissenschaften unternommen.

Im Zentrum der Masterarbeit standen zwei Fragenkomplexe. Einerseits, welche Art von Forschungsdaten bei den untersuchten Fallbeispielen vorhanden sind und welche Bedürfnisse im Kontext der LZA bestehen und andererseits, welche Möglichkeiten die UB Bern hat, um auf die vorgefundenen Bedürfnisse zu reagieren.

Um den ersten Fragekomplex zu untersuchen, wurden teilstrukturierte Interviews⁵ mit Forschenden durchgeführt. Die untersuchten Forschungsprojekte stammen aus den klassischen Disziplinen Geschichte, Sprach- und Literaturwissenschaften der Phil.-Hist.-Fakultät der Universität Bern.⁶ Dabei zielte die Auswahl der wenigen Fallbeispiele nicht auf eine Repräsentativität in statistischem Sinne ab, sondern auf eine konkrete Betrachtung der Fälle in ihrem jeweiligen Umfeld. Für die Interviews wurden vorgängig Leitfäden⁷ und Einwilligungserklärungen⁸ vorbereitet und während den Gesprächen Audioaufnahmen aufgezeichnet. Bei der anschliessenden Verschriftlichung der Aufzeichnungen wurde eine formale Anonymisierung⁹ der teilnehmenden Personen vorgenommen. Der zweite Fragekomplex wurde anhand eines Vergleichs mit anderen Institutionen in der Schweiz und einem abteilungsübergreifenden Austausch mit UB Bern-Mitarbeitenden betrachtet.

Vorliegender Artikel stellt in Kombination mit einzelnen theoretischen Aspekten ausgewählte Ergebnisse dieser Masterarbeit vor.¹⁰

5 Vgl. Werner (2013), S. 130–139.

6 Für Informationen zur Universität Bern vgl. Universität Bern (o. J.).

7 Vgl. Bogner et. al. (2014), S. 27–32; als Orientierung dienen die Beispiele von Bauer et. al. (2015), S. 175–183; Minn/Lemaire (2017), S. 22–31; RatSWD (2016), S. 10–16; SNF (2017).

8 Vgl. Mertes (2013), S. 158; Univation. Institut für Evaluation (2015).

9 Vgl. Bogner et. al. (2014), S. 89.

10 Die vollständige Version der Masterarbeit mit dem Titel «Langzeitarchivierung als forschungsunterstützende Dienstleistung an der Universitätsbibliothek Bern. Bedürfnisse und Möglichkeiten. Eine explorative Pilotstudie im Bereich der Geisteswissenschaften» kann bei der Autorin angefragt werden.

Grundidee der digitalen Langzeitarchivierung

Während es bei der Bestandserhaltung von analogen Medien um den «Substanzerhalt des jeweiligen Objekts»,¹¹ also zum Beispiel um die materielle Erhaltung eines Buches geht, damit die Inhalte lesbar bleiben, bestehen bei der Erhaltung von digitalen Inhalten zusätzliche Herausforderungen.

Bei der LZA von digitalen Medien geht es einerseits um den Erhalt der Daten. Dauerhaft kann dies nur getrennt vom originalen Datenträger und damit unabhängig von dessen unbestimmter Lebensdauer gewährleistet werden.¹² Deshalb müssen die verwendeten Datenträger regelmässig durch neue ersetzt werden. Die Daten sollten redundant – also in mehrfacher Kopie – vorliegen, wobei die Kopien räumlich getrennt und auf unterschiedlichen Datenträgern aufbewahrt werden sollten.¹³

Andererseits geht es auch um den Erhalt der Benutzbarkeit, denn um die erhaltenen Daten auch nutzen zu können, «muss eine aufeinander abgestimmte Kombination aus Hard- und Software zur Verfügung stehen».¹⁴ Dies wird dadurch erschwert, dass sowohl Soft- wie auch Hardware den Entwicklungen des Marktes unterworfen sind und sich dadurch schnell verändern, bzw. durch neue technologische Entwicklungen abgelöst werden.

Zudem müssen unterschiedliche Kontextinformationen mit den Daten, die langzeitarchiviert werden sollen, mitgespeichert werden. Beispielsweise technische Metadaten, um eine automatische Bearbeitung und Verwaltung bei der Umsetzung von Erhaltungsstrategien zu ermöglichen¹⁵ oder bibliographische Metadaten, damit Daten über «grundlegende soziokulturelle Wandlungsprozesse hinaus»¹⁶ nutzbar und interpretierbar bleiben.

Zur Erfüllung einer digitalen LZA stehen unterschiedliche Strategien zur Verfügung.¹⁷ Grundlegend für alle weiteren Schritte ist die Bitstream-Preservation. Diese entspricht der oben beschriebenen Speicherung und Erhaltung des Datenstroms (Bitstream). Darauf aufbauend sind die Strategien der (Format-)Migration und Emulation am verbreitetsten, welche langfristig die Benutzbarkeit der Daten ermöglichen. Bei der Migration werden die Objekte dem neuen Umfeld angepasst, wobei in aktuelle, weit verbreitete, nicht proprietäre Formate migriert wird. Dazu ist langfristig ein personell und finanziell aufwändiger Einsatz notwendig, da Migrationsentscheidungen im laufenden Archivbetrieb von geschulten Personen getroffen werden müssen.

11 Altenhöner /Schrimpf (2015), S. 850.

12 Vgl. Brown (2013), S. 195–197.

13 Vgl. Ullrich (2010), S. 8:4.

14 Altenhöner /Schrimpf (2015), S. 850–851. Sehr einfach und anschaulich ausgedrückt: Alleine das Erhalten von einer CD mit einer Datei stellt nicht sicher, dass die Daten in dieser Datei künftig genutzt werden können. Möglicherweise steht künftig kein CD-Laufwerk zum Einlesen der CD zur Verfügung, oder die Software, die das abgespeicherte Dateiformat darstellen könnte, ist nicht mehr verfügbar.

15 Vgl. Liegmann /Neuroth (2010), S. 1:5.

16 Osswald et. al. (2012), S. 17.

17 Vgl. im Folgenden: Brown (2013), S. 218–228; Funk (2010), S. 8:10; Puhl et. al. (2015), S. 32–33.

Die Emulation geht noch einen Schritt weiter, indem sie für die Objekte deren originales Umfeld simuliert. Dazu werden ganze Softwarepakete zusammen mit den zu archivierenden Dateien in einer Umgebung gespeichert, welche die Rechnerarchitektur der Originalumgebung der Software nachbilden. Dies ist die aufwändigste der genannten Strategien.

Langzeitarchivierende Institutionen können sich dazu am Open Archival Information System (OAIS)-Referenz-Modell orientieren.¹⁸ Es beschreibt, wie digitale Objekte, oft in Verbindung mit der Migrations-Strategie, für eine spezifische künftige Nutzergruppe erhalten werden sollten. Dazu gehören neben einem einheitlichen Vokabular auch ein Datenmodell, ein funktionales Modell sowie ein Überblick zu den Aufgaben, welche die Institution bzw. das Team, das für ein Langzeitarchiv verantwortlich ist, erfüllen muss.

Geisteswissenschaftliche Forschungsdaten und die Zeit

Weil sich geisteswissenschaftliche Forschungsdaten häufig in ihrer Art, Entstehung und Verwendung von nicht-geisteswissenschaftlichen Forschungsdaten unterscheiden, werden sie von gängigen Definitionen für Forschungsdaten nur unzureichend erfasst.¹⁹ Eine eindeutige Definition geisteswissenschaftlicher Forschungsdaten gestaltet sich jedoch schwierig. So bezeichnen Cremer u.a. aufgrund der unterschiedlichen Praktiken und Perspektiven die Forschungsdaten in den Geisteswissenschaften als «vielgestaltiges Mischwesen».²⁰ Eine verbreitete Definition ist diejenige des geisteswissenschaftlichen Infrastrukturprojekts Digital Research Infrastructure for the Arts and Humanities Deutschland (DARIAH-DE):

Unter digitalen geistes- und kulturwissenschaftlichen Forschungsdaten werden innerhalb von DARIAH-DE all jene Quellen/Materialien und Ergebnisse verstanden, die im Kontext einer geistes- und kulturwissenschaftlichen Forschungsfrage gesammelt, erzeugt, beschrieben und/oder ausgewertet werden und in maschinenlesbarer Form zum Zwecke der Archivierung, Zitierbarkeit und zur weiteren Verarbeitung aufbewahrt werden können.²¹

Der SNF empfiehlt, Daten, die während des Forschungsprozesses produziert werden, in öffentlich zugänglichen Repositorien²² für Forschungsdaten zu publizieren. Hierfür stellt der SNF eine Checkliste²³ zur Auswahl fachspezifischer Repositorien zur Verfügung und schlägt alternativ vier fachübergreifende Repositorien vor, darunter auch

18 Hier folgt eine sehr knappe Beschreibung. Ausführliche Informationen finden sich bei: Brübach (2010); Corrado /Sandy (2017), S. 54–63.

19 Vgl. DHd AG Datenzentren (2017), S. 5–7.

20 Cremer et. al. (2018), S. 143.

21 Puhl et. al. (2015), S. 14.

22 Für eine Definition des Begriffs „Repositorium“ vgl. „Forschungsdaten.info (o. J. a).

23 Vgl. Punkt 5.1 bei SNF (o. J. b).

Zenodo.²⁴ Zenodo erfüllt die vom SNF geforderten Kriterien für eine langfristige Sicherung der Daten, übernimmt hingegen keine Verantwortung für deren langfristige Benutzbarkeit und Interpretierbarkeit.²⁵ Dies entspricht einem eher kurz- bis mittelfristigen Zeithorizont für eine mögliche Verwendung der Daten. Dieser Ansatz ist jedoch im Forschungsdatenmanagement²⁶ verbreitet, auch die Deutsche Forschungsgemeinschaft gibt für die Aufbewahrung von Forschungsdaten nur zehn Jahre als Minimum vor.²⁷ Geisteswissenschaftliche Forschungsdaten sind aber «nicht für einen bestimmten Zeitraum (z.B. zehn Jahre) zur Wahrung der Transparenz und Reproduzierbarkeit von Analysen, sondern für einen unbestimmten Zeitraum (im Prinzip, wie bei den Archiven und Bibliotheken: für immer) für eine wiederkehrende Nutzung vorzuhalten.»²⁸ Dies unter anderem auch, weil geisteswissenschaftliche Forschungsdaten oft eine anhaltende wissenschaftliche Relevanz aufweisen, im Gegensatz zu vielen naturwissenschaftlichen Forschungsdaten. Fachübergreifende Repositorien wie Zenodo können für geisteswissenschaftliche Daten genutzt werden, man sollte sich der Problematik der zeitlich limitierten Aufbewahrung jedoch bewusst sein. Dies muss Forschenden, die sich mit dem Aspekt der LZA (bisher) nicht im Detail befasst haben, deutlich kommuniziert werden, da sonst die Gefahr besteht, dass sie davon ausgehen, dass ihre in Repositorien deponierten Daten für immer betreut werden.

In den nächsten drei Kapiteln werden exemplarisch einige Haupteigenschaften der Masterarbeit anhand der drei untersuchten Fallbeispiele aufgeführt.

Langzeitarchivierung von Forschungsdaten – Klassische Missverständnisse

Fallbeispiel 1 (FB1) hat zur Zeit des Interviews²⁹ gerade einen Projektantrag beim SNF eingereicht. Das geplante PhD-Projekt einer Einzelperson ist der Dialektforschung zuzuordnen. Es sollen Interviews mit einer spezifischen Menschengruppe durchgeführt und Tonaufnahmen festgehalten werden, die anschliessend vergleichend mit bestehenden, früheren Aufnahmen von anderen Forschungsgruppen ausgewertet werden. Im Gespräch wurde deutlich, dass FB1 unter LZA eine Speicherung in Form eines Backups versteht. Mit dem Thema LZA kam FB1 bisher nicht in Berührung und geht davon aus, dass man sich bei Projektabschluss darum kümmern wird. Überhaupt

24 Zenodo wird vom CERN betrieben und unter anderen auch von der Europäischen Kommission mitfinanziert. Weiterführende Informationen finden sich auf der Website: Zenodo (o. J. a).

25 Vgl. Zenodo (o. J. b).

26 Für eine Definition des Begriffs „Forschungsdatenmanagement“ vgl. Forschungsdaten.info (o. J. b).

27 Vgl. Deutsche Forschungsgemeinschaft (2015), S. 1 Dies lässt sich u.a. damit begründen, dass die Entwicklungen im Bereich Forschungsdatenmanagement den Fokus nicht primär auf die Geisteswissenschaften gelegt haben.

28 DHd AG Datenzentren (2017), S. 7.

29 Die Informationen zu FB1 stammen aus einem Interview vom 01. Mai 2018.

musste sich FB1 mit vielen Aspekten der geplanten Datenerfassung erstmals im Rahmen des auszufüllenden DMP befassen. Nicht nur die Archivierung, auch die Veröffentlichung der Daten ist nicht vorgesehen, weil für die Daten aufgrund einer rechtlichen Unsicherheit sowie begrenzten Ressourcen keine Anonymisierung vorgesehen ist. Dies ist bedauerlich, da die zu erhebenden Daten für andere Fragestellungen in der Dialektforschung nachgenutzt werden könnten und auch, weil diese Daten einmalig sein werden. Des Weiteren wird bei FB1 beispielhaft deutlich, wie eng die LZA mit anderen Bereichen wie Anonymisierung und projektspezifischen Voraussetzungen zusammenhängen kann und daher immer unter Beachtung von individuellen Faktoren betrachtet werden muss.

Die Feststellungen zur LZA bei FB1 zeigen klassische Missverständnisse auf, die auch Corrado und Sandy hervorheben: «Digital preservation is not only about backups and recovery [, ...] Digital preservation is not only about access [... and] Digital preservation is not an afterthought.»³⁰ Im deutschen Sprachraum ist der Begriff LZA unter Fachexperten zwar verbreitet, er wird jedoch oft missverstanden. So lehnt Alenhöner den Begriff Langzeitarchivierung ab, da er eine «[...] statische Situation suggeriert.»³¹ Stattdessen spricht er von digitaler Langzeitverfügbarkeit, um auf eine aktive Tätigkeit hinzuweisen. Töwe bevorzugt den Begriff digitaler Datenerhalt, denn gerade im Kontext der Forschungsdaten und gegenüber in diesem Gebiet erwarteten Kunden aus der Forschung sei der Begriff LZA missverständlich, da damit fälschlicherweise eine nach Projektabschluss passiv stattfindende Abgabe der Daten zur Archivierung vermittelt werden könnte.³² In vorliegendem Artikel wird jedoch der Begriff LZA verwendet, da er an der UB Bern vertreten und in der deutschsprachigen Literatur weit verbreitet ist. Anstatt den Begriff durch neue Konstruktionen zu ersetzen, wird es als zentral erachtet, auf Bedeutung und Wichtigkeit der digitalen LZA aufmerksam zu machen, mit dem Ziel, sowohl die Daten zu erhalten als auch für deren Benutzbarkeit zu sorgen.

Digitale Edition, NIE-INE und DaSCH

Fallbeispiel 2 (FB2) ist ein umfassendes Editionsprojekt aus dem Fachbereich der Literaturwissenschaften. Parallel und auf Basis derselben Daten entstehen eine mehr als 60-bändige Buchedition und künftig auch eine digitale Edition. Die digitale Edition wird gegenüber der Printedition einen Mehrwert aufweisen, unter anderem weil zielgruppenspezifische Ansichten derselben Daten ermöglicht werden sollen. Die Ressourcenintensivität rechtfertigt sich dadurch, dass die breite Öffentlichkeit Zugang zu

30 Corrado /Sandy (2017), S. 4–5.

31 Alenhöner /Schrimpf (2015), S. 853.

32 Vgl. Töwe (2015), S. 167.

den Ergebnissen der editorischen Arbeit des Forschungsprojekts haben wird, das sich mit schweizerischem Kulturerbe befasst. Das an einer dafür eingerichteten Forschungsstelle angesiedelte Projekt steht zur Zeit des Interviews etwa in der Mitte der vorgesehenen 34-jährigen Projektdauer, hat in den letzten Jahren erfolgreich auf eine digitale Arbeitsumgebung gewechselt und konzipiert aktuell die digitale Edition.³³

Eine digitale Edition kann als komplexe digitale Ressource bezeichnet werden. Während geisteswissenschaftliche Forschungsdaten «mehrerlich voneinander unabhängige Einzeldateien sind, verfügen digitale Editionen und Datenbank[en] meist über einen die Summe der Einzeldateien überschreitenden und häufig auch funktionalen Mehrwert».³⁴ Häufig liegen zusätzlich zu den Daten auch archivierungswürdige Systeme³⁵ vor. Erhaltenswerte Forschungsergebnisse sind in diesen Fällen dann «mehr als nur Forschungsdaten im engeren Sinne: Auch Präsentationsumgebungen und Applikationen, die Daten interpretieren, Möglichkeiten zur Recherche und Verknüpfung schaffen sowie Services zur Nutzung und Weiterverarbeitung anbieten [...]»³⁶ zählen dazu.

Für die in solchen Fällen oft kompliziertere LZA können Datenzentren genutzt werden. Die Arbeitsgruppe Datenzentren der Digital Humanities im deutschsprachigen Raum bezeichnet geisteswissenschaftliche Datenzentren als «lebende Archive».³⁷ Dies, weil sie wie Archive institutionell auf Dauer angelegt sind und auf eine nachhaltige Verfügbarkeit der Inhalte abzielen, jedoch gleichzeitig auch permanenten Online-Zugang zu einzelnen Objekten und den Betrieb laufender generischer oder projektspezifischer Softwaresysteme gewährleisten, sowie aktives Forschungsdatenmanagement während und nach Projektende von Forschungsprojekten betreiben. Damit wird im Unterschied zu (den meisten) Forschungsdaten-Repositoryen mehr als nur Bitstream-Preservation angeboten, da die Daten hier langfristig gepflegt werden.

FB2 ist genau ein solcher Fall, in dem nicht nur die Daten erhaltenswert sind, sondern auch das System. Weil sich die technische Beratung nicht unbedingt innerhalb des gewohnten Netzwerks eines Geisteswissenschaftlers finden lässt, hat FB2 schon früh nach Unterstützung gesucht. Da sich diese zu Projektbeginn in der Schweiz nicht finden liess, holte sich FB2 Beratung bei mehreren Kompetenzzentren in Deutschland. Mittlerweile steht fest, dass künftig mit dem Projekt Nationale Infrastruktur für Editionen (NIE-INE) in der Schweiz zusammengearbeitet werden soll.

Bei NIE-INE wird eine Arbeitsoberfläche mit modularen Funktionen spezifisch für digitale Editionen erschaffen.³⁸ Zentrale Ziele von NIE-INE sind die

33 Die Informationen zu FB2 stammen aus einem Interview vom 17. Mai 2018.

34 Andorfer (2015), S. 22.

35 Sahle /Kronenwett (2013), S. 83.

36 Vgl. DHd AG Datenzentren (2017), S. 8.

37 Ebd., S. 10.

38 NIE-INE (o. J.).

langfristige Sicherung, Verfügbarkeit und Publikation der digitalen Forschungsergebnisse sowie deren zugrunde liegenden Forschungsdaten. Dazu wird die Infrastruktur vom Data and Service Center for the Humanities (DaSCH)³⁹ genutzt. Zu den primären Zielen von NIE-INE gehören auch die Beratung und Betreuung der Editionsprojekte sowie allgemeine Services, beispielsweise für technische Anpassungen. Koordiniert wird das Projekt vom an der Universität Basel angesiedelten Forum für Edition und Erschließung (FEE).⁴⁰ NIE-INE ist zurzeit (Stand 2018) noch in der Projektphase.

Das DaSCH ist seit 2017 ein dauerhaftes Unternehmen der Schweizerischen Akademie der Geistes- und Sozialwissenschaften und damit ein nationales Datenzentrum.⁴¹ Es gewährleistet unter anderem die langfristige Sicherung und (Nach-)Nutzbarkeit von qualitativen geisteswissenschaftlichen Forschungsdaten. Dies beinhaltet auch das Hosting von für die Geisteswissenschaften relevanten Daten.⁴²

Technisch besteht das DaSCH aus einer offenen, modularen, erweiterbaren und flexiblen Plattform, die auf industriellen Standards wie Resource Description Framework basiert, worauf die Daten gespeichert und verwaltet werden. Diese Plattform wird ergänzt durch eine generische, virtuelle Forschungsumgebung, in der die Daten bearbeitet werden können; auf Wunsch können auch projektspezifische Zugänge eingerichtet werden.⁴³

Für die LZA von komplexen Ressourcen, die nicht nur aus Daten, sondern auch aus Systemen bestehen und auf die eine Zugriffsmöglichkeit bestehen soll, wird vom DaSCH ein Keep-Alive-Archiving-Ansatz in Kombination mit einer Archivierung nach dem OAIS-Referenzmodell verfolgt.⁴⁴ Mithilfe von Zeitstempeln können auch Projekte unterstützt werden, die inhaltlich noch verändert werden, denn so bleiben sie zitierfähig.⁴⁵ Sobald sich die Technologie weiterentwickelt, wird das ganze System migriert.

Das DaSCH ist speziell, aber nicht ausschliesslich für komplexe Daten geeignet. Grundsätzlich können hier alle geisteswissenschaftlichen Daten sowohl aus neu startenden als auch bereits laufenden oder abgeschlossenen Projekten betreut werden. Wobei angefügt werden muss, dass bei bereits laufenden oder abgeschlossenen Projekten allenfalls ein relativ grosser Aufwand entstehen kann, falls das Datenmodell oder die Datenstruktur angepasst werden müssen.

39 Das DaSCH wird im nächsten Abschnitt ausgeführt.

40 FEE (o. J.).

41 Vgl. SAGW (o. J. b).

42 Vgl. DaSCH (o. J.).

43 Vgl. Rosenthaler et. al. (2015), S. 17–19.

44 Vgl. Ebd., S. 16.

45 Auf Anfrage wurde dies am 12. Juli 2018 von der DaSCH-Leitung per E-Mail bestätigt.

Relationale Datenbank aus fortlaufendem Projekt

Fallbeispiel 3 (FB3) bearbeitet seit 17 Jahren eine umfassende Datenbank mit Einträgen zu einer wissenschaftlich relevanten Gruppe von Personen. Die Personen werden aufgrund von Universitätsmatrikellisten erfasst und bei der weiteren Bearbeitung mit Information zu Herkunft, verwandtschaftlichen Bindungen, Lebensweg, beruflichen Stationen, verfassten Werken und Weiterem angereichert. Ein Webzugriff ermöglicht unter anderem die dynamische Erzeugung von Visualisierungen der täglich aktualisierten Daten. Für dieses Projekt aus dem Fachbereich Geschichte ist vorerst kein Abschluss geplant, die bestehenden Daten werden weiter angereichert.⁴⁶

FB3 spielt schon seit einigen Jahren mit dem Gedanken, seine Daten bei einer auf Dauer ausgelegten Institution wie einer Bibliothek unterzubringen. Es wurden unterschiedliche Überlegungen zum Thema LZA gemacht, mit dem Fazit, dass die Datenbank idealerweise am laufenden Betrieb zu halten ist. Dazu sind FB3 jedoch keine passenden Angebote bekannt. Eine Schwierigkeit ist, dass ähnlich wie bei FB2 komplexere Daten vorliegen, wobei zusätzlich zu den Daten auch ein System vorhanden ist, welches ebenso langzeitarchiviert werden muss.

Speziell für die Archivierung von relationalen Datenbanken wurde vom Schweizerischen Bundesarchiv SIARD (Software independent archiving of relational databases)⁴⁷ entwickelt. Aber SIARD deckt nicht alle Bedürfnisse von FB3 ab. Eine Problematik stellt Folgendes dar: Um die Daten anzusehen oder zu nutzen, muss das ganze Datenset aus dem Archiv geholt und mit SIARD-Suite in ein funktionierendes Datenbankmanagementsystem überführt werden – es ist nicht möglich, bloss einen kurzen Blick auf einen einzelnen Datensatz zu werfen.⁴⁸ So eignet sich SIARD wirklich nur für abgeschlossene Datenbanken, die quasi eingefroren werden können. Das macht nur dann Sinn, wenn sie nicht mehr verändert werden und auf die Daten nicht mehr regelmässig zugegriffen werden soll. Das ist bei FB3 jedoch nicht der Fall. Trotzdem wird gewünscht, das langjährige, ressourcenintensive Projekt bald absichern zu können, wenn auch nur ein gewisser Stand der Daten. Zudem werden die Darstellungsmöglichkeiten der Daten über den Webzugriff als zentrales Element von FB3 betrachtet, damit die Datenbank als Forschungsinstrument genutzt werden kann. Dies soll auch künftig möglich sein.

Wie von FB3 selbst gesagt, wäre es die beste Möglichkeit, die Datenbank und den Webzugriff am laufenden Betrieb zu halten. Dazu könnte das DaSCH in Betracht gezogen werden. Denn es ist, wie im vorangehenden Kapitel ausgeführt, spezialisiert auf komplexere geisteswissenschaftliche Daten, auf die auch nach einer Abgabe ans DaSCH regelmässig zugegriffen werden kann und bietet zudem Lösungen an, wenn

46 Die Informationen zu FB3 stammen aus einem Interview vom 27. April 2018.

47 Vgl. Schweizerisches Bundesarchiv (2018).

48 Vgl. Rosenthaler et. al. (2015), S. 16.

die Inhalte noch verändert werden sollen. FB3 hat bisher keinen Kontakt zum DaSCH aufgenommen, da eine Problematik darin besteht, dass die vorhandene relationale Datenbank durch ihr historisches Wachstum in Datenstruktur und Datenmodell sehr komplex geworden ist und stark verändert werden müsste, um mit der Infrastruktur des DaSCH archiviert werden zu können. Dieser Aufwand wird als zu hoch eingeschätzt.

Weil die Forschungsergebnisse von FB3 jedoch einen hohen wissenschaftlichen Wert mit sich bringen, indem die Datenbank als Forschungsinstrument funktioniert, somit als Grundlage für weitere Forschung dient und zudem einzigartige Daten nutzbar machen kann, rechtfertigt sich die Ressourcenintensivität des Forschungsprojekts. Diese Punkte befürworten eine LZA, weshalb FB3 weiter nach einer zufriedenstellenden Lösung sucht.

Langzeitarchivierung an der UB Bern bisher

Anhand der drei untersuchten Fallbeispiele wurden klassische Missverständnisse bei der LZA von geisteswissenschaftlichen Forschungsdaten vorgeführt, Schwierigkeiten und Lösungsansätze bei der LZA von komplexen digitalen Ressourcen (digitale Edition sowie relationale Datenbank) aufgezeigt und die grundsätzliche Frage der Langzeitarchivierung bei fortlaufenden Forschungsprojekten (am Leben halten oder «Einfrieren») besprochen.

Ein Teil der Strategie⁴⁹ für die LZA an der UB Bern sieht für 2017-2020 vor, dass eine digitale Langzeitarchivierungsstrategie implementiert und ein Langzeitarchiv eingerichtet wird. Einiges ist bereits erreicht: Die Abteilung LZA kümmert sich um die LZA ihrer E-Books und E-Journals, indem sie sich an kooperativen Projekten beteiligt. Zudem wurde mit Bern Digital Archive (BerDA) eine Infrastruktur für die LZA der UB Bern aufgebaut. Sie ist auf die LZA der auf e-rara.ch verfügbaren Digitalisaten der UB Bern ausgerichtet und ist daher vorerst ohne externe Zugriffsmöglichkeit in Betrieb genommen worden.

Die Abteilung LZA ist aktuell noch nicht auf Unterstützung für Forschende ausgerichtet, da bisher unklar war, welche Bedürfnisse vorhanden sind. Weil die Strategie der UB Bern einen von fünf Schwerpunkten auf die Unterstützung von Bildung, Forschung und Lehre setzt, werden nun bestehende Angebote ausgebaut. Dies betrifft unter anderem die Bereiche elektronisches Publizieren und Forschungsdatenmanagement, aber auch die LZA.⁵⁰

49 Vgl. Universitätsbibliothek Bern (2017), S. 4; Für Informationen zur Universitätsbibliothek Bern vgl. Universitätsbibliothek Bern (o. J. a).

50 Die aktuellen Dienstleistungen für Forschende an der UB Bern werden auf der Website übersichtlich dargestellt, vgl. Universitätsbibliothek Bern (o. J. b).

Mögliche Weiterentwicklung der Langzeitarchivierung an der UB Bern

Auf Grundlage der durch Interviews mit unterschiedlichen Fallbeispielen erhobenen Ergebnisse können für die mögliche Weiterentwicklung der LZA an der UB Bern folgende Punkte festgestellt werden:

- Der Informationsbedarf der betrachteten Fallbeispiele ist sehr unterschiedlich. Dies, weil die Projekte sowohl in Bezug auf den Projektstand als auch in Bezug auf die Überlegungen zur LZA der Daten unterschiedlich fortgeschritten sind.
- FB2 suchte sich seine Informationen im Ausland zusammen, da zu Projektbeginn vor Ort kein Angebot bekannt war – das kann auch bei anderen, ähnlichen Projekten der Fall sein. Die UB Bern als Ansprechpartner im Bereich LZA in Betracht zu ziehen, ist (noch) nicht verbreitet. Sie muss sich bei einem Ausbau der Angebote in dieser Rolle von Anfang an bekannt machen.
- Auch die inhaltliche Breite und Tiefe der festgestellten Informationsbedürfnisse ist sehr unterschiedlich. Bei FB1 war LZA noch gar nicht bekannt, Bei FB2 und FB3 war das Thema zwar mittlerweile bekannt, es wurde nun aber eher ein Überblick zu bestehenden Angeboten von unterschiedlichen Anbietern und zu komplexeren Lösungen gewünscht.
- Alle drei Fallbeispiele äusserten den Wunsch nach einer ersten Anlaufstelle, die nicht nur bei Fragen der LZA, sondern auch bei Themen wie Digitalisierung, Dateibenennung, Anonymisierung, Datenmanagement usw. weiterhelfen oder an Experten verweisen kann. Dieses Bedürfnis deckt sich mit Ergebnissen aus Interviews, durchgeführt von einer internen Arbeitsgruppe, die sich mit dem Ausbau des digitalen Dienstleistungsangebots der UB Bern befasst. Mittlerweile wird bei einer weiteren Arbeitsgruppe an einem übersichtlichen Angebot und der Vermittlung von forschungsunterstützenden Dienstleistungen im Bereich der Geisteswissenschaften gearbeitet, wobei ebenfalls die Umsetzung eines Angebots, das einer ersten Anlaufstelle entspricht, verfolgt wird.
- Auf Grundlage der erhobenen Bedürfnisse könnte das Angebot in Bezug auf Infrastruktur für die LZA für Forschende folgendermassen ausgerichtet werden: BerDA kann ohne Oberflächensystem für Forschungsprojekte genutzt werden, die wie FB3 einen eigenen Webzugang betreuen und parallel dazu ihre Daten langzeitarchivieren möchten. In diesem Sinne wird FB3 zurzeit als mögliches Pilotprojekt in Betracht gezogen. Auch ein Ausbau von BerDA, der einen externen Zugriff auf die Daten ermöglicht, könnte in Betracht gezogen werden. Zudem kann auf externe Angebote wie das DaSCH verwiesen werden.
- An der UB Bern befindet sich bei der Abteilung Forschungsdatenmanagement ein institutionelles Repositorium für Forschungsdaten im Aufbau. Dieses soll den Forschenden ermöglichen, die Forderungen des SNF umzusetzen, sofern für ihre Daten keine fachspezifischen Repositorien zur Verfügung stehen.

Komplexe Daten wie Datenbanken und digitale Editionen werden darauf jedoch nicht abgelegt werden können. Da ein generisches Repositorium für die ganze Universität vorgesehen ist, welches somit nicht spezifisch auf geisteswissenschaftlichen Daten ausgerichtet sein wird, wird es sich an einem kurz- bis mittelfristigen Zeithorizont von 10 Jahren als Minimum für die garantierte Aufbewahrung orientieren, ähnlich wie Zenodo. BerDA ist zwar langfristiger ausgelegt, bietet jedoch bisher keine externe Zugriffsmöglichkeit, was den Forderungen der forschungsfördernden Institutionen nicht gerecht wird. Hier könnte in Betracht gezogen werden, eine Schnittstelle zwischen dem künftigen institutionellen Repositorium für Forschungsdaten und BerDA einzurichten, damit sich die Angebote noch einfacher ergänzen könnten.

- Auch die Beratung würde idealerweise Themen des Forschungsdatenmanagements und der LZA kombinieren. Einerseits, weil Forschende aktuell vor allem durch den Druck, den forschungsfördernde Institutionen wie der SNF mit den auszufüllenden DMP's ausüben, Unterstützung im Forschungsdatenmanagement suchen. Andererseits auch, weil die LZA in den gesamten Lebenszyklus⁵¹ von Forschungsdaten eingebunden ist und nicht separat betrachtet werden sollte. Da die Abteilung LZA sich bisher mit den eigenen Digitalisaten der UB Bern befasst hat, war kein Austausch mit der Abteilung Forschungsdatenmanagement notwendig. Dies ändert sich jedoch, sobald es um die LZA von Forschungsdaten geht, weil diese nicht losgelöst von ihrem Lebenszyklus betrachtet werden können. So ist es unabdingbar, dass die beiden Abteilungen enger zusammenarbeiten.

Vorliegender Artikel hat unterschiedliche Bedürfnisse wie auch erste Schritte zur Weiterentwicklung skizziert, um das bestehende Angebot zur LZA der UB Bern auf Forschungsdaten auszurichten. Welche von diesen Vorschlägen künftig umgesetzt und welche weiteren Entwicklungen vorgenommen werden (können), wird erst die Zukunft zeigen. Es ist jedoch unabdingbar für die UB Bern, das Bewusstsein der Notwendigkeit der LZA von Forschungsdaten zu fördern sowie entsprechende Dienstleistungen anzubieten. Denn die rasanten technologischen (Weiter-)Entwicklungen, die im Rahmen der Digitalisierung fortschreiten, verursachen ein Kaleidoskop an kaum greifbaren, sich ständig verändernden Möglichkeiten – auch für die technologische Umsetzung von Forschungsprojekten. So unternimmt die UB Bern einen notwendigen ersten Schritt, um dem Risiko des Verlusts der Benutzung oder sogar der Daten an sich entgegenzuwirken.

51 DARIAH-DE entwickelt spezifisch für die Geisteswissenschaften ein Referenzmodell eines Research Data LifeCycle, Vgl. Puhl et. al. (2015), S. 7.

Abkürzungen

BerDA = Bern Digital Archive

DARIAH-DE = Digital Research Infrastructure for the Arts and Humanities Deutschland

DaSCH = Data and Service Center for the Humanities

DMP = Data Management Plan

FB1, FB2, FB3 = Fallbeispiel 1, Fallbeispiel 2, Fallbeispiel 3

LZA = Langzeitarchivierung, synonym zu digitale Langzeitarchivierung

NIE-INE = Nationale Infrastruktur für Editionen

SIARD = Software independent archiving of relational databases

UB Bern = Universitätsbibliothek Bern

Bibliografie

Altenhöner, Reinhard/Schrimpf, Sabine (2015): Bestandserhaltung und Langzeitverfügbarkeit digitaler Ressourcen. Strategie, Organisation und Techniken, in: Rolf Griebel et. al. (Hg.): *Praxishandbuch Bibliotheksmanagement* (Bd. 2), Berlin, S. 850–872.

Andorfer, Peter (2015): Forschungsdaten in den (digitalen) Geisteswissenschaften. Versuch einer Konkretisierung, Göttingen, <http://webdoc.sub.gwdg.de/pub/mon/dariah-de/dwp-2015-14.pdf>, Zugriff: 13.07.2018.

Bauer, Bruno et. al. (2015): Forschende und ihre Daten. Ergebnisse einer österreichweiten Befragung – Report 2015. Version 1.1, <https://zenodo.org/record/32037#.WoyMLOciE2w>, Zugriff: 15.07.2019.

Bogner, Alexander et. al. (2014): Interviews mit Experten. Eine praxisorientierte Einführung, Wiesbaden.

Brown, Adrian (2013): *Practical digital preservation. A how-to guide for organizations of any size*, London.

Brübach, Nils (2010): Das Referenzmodell OAIS, in: Heike Neuroth et. al. (Hg.): *Nestor-Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*, Boizenburg, Version 2.3, S. 4:3-4:14.

Corrado, Edward M./Sandy, Heather M. (2017): *Digital preservation for libraries, archives, and museums* (2. Aufl.), Lanham.

Cremer, Fabian et. al. (2018): Der Chimäre auf der Spur. Forschungsdaten in den Geisteswissenschaften, in: *O-bib. Das offene Bibliotheksjournal* 5 (2), S. 142–162.

DaSCH (o. J.): DaSCH. Services, <http://dasch.swiss/services/>, Zugriff: 22.07.2018.

Deutsche Forschungsgemeinschaft (2015): Leitlinien zum Umgang mit Forschungsdaten, http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf, Zugriff: 14.07.2018.

DHD AG Datenzentren (2017): Geisteswissenschaftliche Datenzentren im deutschsprachigen Raum. Grundsatzpapier zur Sicherung der langfristigen Verfügbarkeit von Forschungsdaten, <https://doi.org/10.5281/zenodo.1134760>, Zugriff: 13.07.2018.

FEE (o. J.): Nationale Infrastruktur für Editionen. Infrastructure nationale pour les éditions (NIE-INE), http://www.fee.unibas.ch/nie_ine.html, Zugriff: 14.07.2018.

Forschungsdaten.info (o. J. a): Glossar. Repositorium, <https://www.forschungsdaten.info/support/glossar/#c269860>, ugriff: 14.07.2018.

——— (o. J. b): Glossar. Forschungsdatenmanagement, <https://www.forschungsdaten.info/support/glossar/#c269836>, Zugriff: 17.07.2019.

Funk, Stefan E. (2010): Migration, in: Heike Neuroth et. al. (Hg.): *Nestor-Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*, Boizenburg, Version 2.3, S. 8:10-8:15.

Liegmann, Hans/Neuroth, Heike (2010): Einführung, in: Heike Neuroth et. al. (Hg.): *Nestor-Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*, Boizenburg, Version 2.3, S. 1:1-1:10.

Mertes, Nathalie (2013): Fallstudien, in: Konrad Umlauf et. al. (Hg.): *Handbuch Methoden der Bibliotheks- und Informationswissenschaft*, Berlin, S. 152–167.

Minn, Gisela/Lemaire, Marina (2017): Forschungsdatenmanagement in den Geisteswissenschaften. Eine Planungshilfe für die Erarbeitung eines digitalen Forschungskonzepts und die Erstellung eines Datenmanagementplans, Trier, <http://ubt.opus.hbz-nrw.de/volltexte/2017/1071/>, Zugriff: 14.07.2018.

NIE-INE, Nationale Infrastruktur für Editionen (o. J.): Über NIE-INE, <https://www.nie-ine.ch/steckbrief-de>, Zugriff: 15.07.2019.

Osswald, Achim et. al. (2012): Langzeitarchivierung von Forschungsdaten. Einführende Überlegungen, in: Heike Neuroth et. al. (Hg.): *Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme*, Boizenburg, S. 13–21.

Puhl, Johanna et. al. (2015): Diskussion und Definition eines Research Data LifeCycle für die digitalen Geisteswissenschaften, Göttingen, <http://webdoc.sub.gwdg.de/pub/mon/dariah-de/dwp-2015-11.pdf>, Zugriff: 14.07.2018.

RatSWD (Hg.) (2016): Forschungsdatenmanagement in den Sozial-, Verhaltens- und Wirtschaftswissenschaften. Orientierungshilfen für die Beantragung und Begutachtung datengenerierender und datennutzender Forschungsprojekte (2. Aufl.), Berlin, <https://doi.org/10.17620/02671.7>, Zugriff: 14.07.2018.

Rosenthaler, Lukas et. al. (2015): Final report for the pilot project „Data and Service Center for the Humanities“ (Dasch), <https://doi.org/10.5281/zenodo.822918>, Zugriff: 14.07.2018.

SAGW (o. J. a): Glossar „Open Access“, https://sagw.ch/fileadmin/user_upload/Glossar_Open_Access.pdf, Zugriff: 15.07.2019.

——— (o. J. b): Data and Service Center for the Humanities (DaSCH), <https://sagw.ch/dasch/>, Zugriff: 15.07.2019.

Sahle, Patrick/Kronenwett, Simone (2013): Jenseits der Daten. Überlegungen zu Datenzentren für die Geisteswissenschaften am Beispiel des Kölner „Data Center for the Humanities“, in: *LIBREAS. Library Ideas* 23, S. 76–96.

Schweizerisches Bundesarchiv (2018): SIARD Suite, <https://www.bar.admin.ch/bar/de/home/archivierung/tools---hilfsmittel/siard-suite.html>, Zugriff: 22.07.2018.

SNF (o. J. a): Open Research Data, http://www.snf.ch/de/derSnf/forschungspolitische_positionen/open_research_data/Seiten/default.aspx, Zugriff: 15.07.2019.

——— (o. J. b): Data Management Plan (DMP). Leitlinien für Forschende, http://www.snf.ch/de/derSnf/forschungspolitische_positionen/open_research_data/Seiten/data-management-plan-dmp-leitlinien-fuer-forschende.aspx, Zugriff: 14.07.2018.

——— (2017): Data management plan - mySNF Formular, http://www.snf.ch/SiteCollectionDocuments/DMP_content_mySNF-form_de.pdf, Zugriff: 15.07.2019.

Töwe, Matthias (2015): Von Forschungsdaten zu e-journals und zurück. Der Weg zum digitalen Datenerhalt an der ETH-Bibliothek, in: Rafael Ball/Stefan Wiederkehr (Hg.): *Vernetztes Wissen. Online. Die Bibliothek als Managementaufgabe*, Berlin, S. 159–171.

Ullrich, Datmar (2010): Bitstream Preservation, in: Heike Neuroth et. al. (Hg.): *Nestor-Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*, Boizenburg, Version 2.3, S. 8:3-8:9.

Univation. Institut für Evaluation (2015): Informiertes Einverständnis. Eval-Wiki: Glossar der Evaluation., https://eval-wiki.org/w_glossar/index.php?title=Informiertes_Einverst%C3%A4ndnis&oldid=1080, Zugriff: 14.07.2018.

Universität Bern (o. J.): Universität Bern. Website, www.unibe.ch, Zugriff: 17.07.2019.

Universitätsbibliothek Bern (o. J. a): Universitätsbibliothek Bern. Website, www.ub.unibe.ch, Zugriff: 17.07.2019.

——— (o. J. b): Open Science, http://www.unibe.ch/universitaet/dienstleistungen/universitaetsbibliothek/service/open_science/index_ger.html, Zugriff: 22.07.2018.

——— (2017): Strategie 2017-2020, http://www.unibe.ch/unibe/portal/content/e809/e962/e963/e6382/e6386/e552940/Strategie_A5_Web_ger.pdf, Zugriff: 14.07.2018.

Werner, Petra (2013): Qualitative Befragungen, in: Konrad Umlauf et. al. (Hg.): *Handbuch Methoden der Bibliotheks- und Informationswissenschaft*, Berlin, S. 128–151.

Zenodo (o. J. a): Zenodo about. Infrastructure, <http://about.zenodo.org/infrastructure/>, Zugriff: 22.07.2018.

——— (o. J. b): General policies, <http://about.zenodo.org/policies/>, Zugriff: 15.07.2019.