

GOEDOC – Dokumenten- und Publikationsserver der Georg-August-Universität Göttingen

2015

Forschen und Forschungsdaten in den Geisteswissenschaften – Zwischenbericht einer Interviewreihe

Peter Andorfer (Herzog August Bibliothek Wolfenbüttel)

DARIAH-DE Working Papers

Nr. 10

Andorfer, Peter: Forschen und Forschungsdaten : Zwischenbericht einer Interviewreihe
Göttingen : GOEDOC, Dokumenten- und Publikationsserver der Georg-August-Universität, 2015
(DARIAH-DE working papers 10)

Verfügbar:

PURL: <http://resolver.sub.uni-goettingen.de/purl/?dariah-2015-3>

URN: <http://nbn-resolving.de/urn:nbn:de:gbv:7-dariah-2015-3-8>

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)



Bibliographische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

Erschienen in der Reihe
DARIAH-DE working papers

ISSN: 2198-4670

Herausgeber der Reihe
DARIAH-DE, Niedersächsische Staats- und Universitätsbibliothek

Mirjam Blümm, Thomas Kollatz, Stefan Schmunk und Christof Schöch

Abstract: Im Rahmen der an der Herzog August Bibliothek Wolfenbüttel durchgeführten Interviewreihe: „Forschen und Forschungsdaten in den Geisteswissenschaften“ wurden GeisteswissenschaftlerInnen und StipendiatInnen zu ihrer Forschungspraxis und ihrem Umgang mit Forschungsdaten befragt. Vorliegender Bericht präsentiert erste Ergebnisse (Stand Jänner 2015) der auf <http://digital-archiv.at:8081/exist/apps/DARIAH-Collection/pages/InterviewAuswertung.html> „in Echtzeit“ erfolgenden Auswertung.

Keywords: Geisteswissenschaften; Digitale Geisteswissenschaften; Forschungsdaten; Umfrage
Humanities; Digital Humanities; research data; questionnaire

Bibliographische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

Erschienen in der Reihe
DARIAH-DE working papers

ISSN: 2198-4670

Herausgeber der Reihe
DARIAH-DE, Niedersächsische Staats- und Universitätsbibliothek

Mirjam Blümm, Thomas Kollatz, Stefan Schmunk und Christof Schöch

Abstract: Im Rahmen der an der Herzog August Bibliothek Wolfenbüttel durchgeführten Interviewreihe: „Forschen und Forschungsdaten in den Geisteswissenschaften“ wurden GeisteswissenschaftlerInnen und StipendiatInnen zu ihrer Forschungspraxis und ihrem Umgang mit Forschungsdaten befragt. Vorliegender Bericht präsentiert erste Ergebnisse (Stand Jänner 2015) der auf <http://digital-archiv.at:8081/exist/apps/DARIAH-Collection/pages/InterviewAuswertung.html> „in Echtzeit“ erfolgenden Auswertung.

Keywords: Geisteswissenschaften; Digitale Geisteswissenschaften; Forschungsdaten; Umfrage; Humanities; Digital Humanities; research data; questionnaire

Forschen und Forschungsdaten in den Geisteswissenschaften. Zwischenbericht einer Interviewreihe

Peter Andorfer
(Herzog August Bibliothek Wolfenbüttel)



Peter Andorfer: „Forschen und Forschungsdaten in den Geisteswissenschaften.
Zwischenbericht einer Interviewreihe“. *DARIAH-DE Working Papers* Nr. 10.
Göttingen: DARIAH-DE, 2015. URN: urn:nbn:de:gbv:7-dariah-2015-3-8.

Dieser Beitrag erscheint unter der
Lizenz [Creative-Commons Attribution 4.0](https://creativecommons.org/licenses/by/4.0/) (CC-BY).

Die *DARIAH-DE Working Papers* werden von Mirjam Blümm,
Thomas Kollatz, Stefan Schmunk und Christof Schöch
herausgegeben.



Abstract

Im Rahmen der an der Herzog August Bibliothek Wolfenbüttel durchgeführten Interviewreihe: „Forschen und Forschungsdaten in den Geisteswissenschaften“ wurden GeistwissenschaftlerInnen und StipendiatInnen zu ihrer Forschungspraxis und ihrem Umgang mit Forschungsdaten befragt. Vorliegender Bericht präsentiert erste Ergebnisse (Stand Jänner 2015) der auf <http://digital-archiv.at:8081/exist/apps/DARIAH-Collection/pages/InterviewAuswertung.html> „in Echtzeit“ erfolgenden Auswertung.

Inhaltsverzeichnis

1 Einleitung	4
1.1 Motivation und Projekthintergrund.....	4
1.2 Forschungsstand.....	5
1.3 Besonderheiten dieser Interviewreihe.....	7
1.4 Technische Umsetzung.....	7
2 Ergebnisse I: Programme, Dateiformate & Arbeitsschritte ...	8
2.1 MS-Word und/oder Citavi, Zotero & Co.....	8
2.2 Datenmanagement: Windows/Mac-Explorer oder Citavi, Zotero & Co	9
2.3 Typische Arbeitsschritte: Bibliographieren, Notieren & Schreiben.....	9
2.4 Spezifische Arbeitsschritte.....	10
3 Ergebnisse II: Forschungsdaten in den Geistes- und Kulturwissenschaften	10
3.1 Forschungsdaten zum Zwecke der Nachvollziehbarkeit.....	10
3.2 Forschungsdaten zum Zwecke der Nachnutzung.....	11
4 Abschluss und Ausblick	12
5 Literaturverzeichnis	13
6 Links	14

1 Einleitung

Wie forschen Geistes- und KulturwissenschaftlerInnen, welche konkreten Tätigkeiten (Bibliographieren, Transkribieren, Schreiben, ...) üben sie aus und welche Programme (MS-Word, Zotero, ...) und Dateiformate (docx, pdf, ...) verwenden sie? Welche (Forschungs)Daten fallen im Laufe eines geistes- und kulturwissenschaftlichen Projektes an, wie werden diese Daten während des Forschens verwaltet, wie werden sie gespeichert und was passiert mit den Daten nach Abschluss des Projektes? Diese und weitere Punkte werden in einer seit Herbst 2014 laufenden Interviewserie mit WissenschaftlerInnen, die mit den Beständen an der Herzog August Bibliothek in Wolfenbüttel arbeiten, abgefragt. Das vorliegende Paper evaluiert einerseits Potentiale und Mängel dieser Interviewreihe und arbeitet Verbesserungsvorschläge aus, benennt andererseits aber auch Ergebnisse und Trends. Zuvor werden jedoch noch der konkrete Hintergrund des Projektes beschrieben, thematisch-methodisch ähnliche Projekte bzw. Interviews und Umfragen vorgestellt, um daran anschließend die Besonderheiten dieser Interviewreihe aufzeigen zu können und es wird die technische Umsetzung der Dokumentation der Interviews erläutert.

1.1 Motivation und Projekthintergrund

Die hier beschriebene Interviewreihe wird im Rahmen des Projektes DARIAH-DE durchgeführt.¹ Die Herzog August Bibliothek² ist als einer von 20 Projektpartnern³ in Cluster 4 „Wissenschaftliche Sammlungen und Forschungsdaten“⁴ des insgesamt sechs Cluster umfassenden Projekts engagiert. Ziele dieses Clusters sind unter anderem die Erarbeitung weitgehend generischer Konzepte und Definitionen der Begriffe ‚Wissenschaftliche Sammlungen‘ und ‚Forschungsdaten‘, sowie die Ausarbeitung von Nutzungsszenarien und Use-Cases, entsprechend der konkreten Bedürfnisse seitens der einzelnen Fachwissenschaften. Neben der Herzog August Bibliothek sind unter der Leitung der Niedersächsischen Staats- und Universitätsbibliothek Göttingen⁵ noch die Berlin Brandenburgische Akademie der Wissenschaften,⁶ das Deutsche Archäologische Institut⁷ und das Leibniz-Institut für Europäische Geschichte⁸ in Cluster 4 involviert. Auf Basis der zu erarbeitenden Definitionen und Use-Cases werden Tools und Services zur Beschreibung, Archivierung und Veröffentlichung von wissenschaftlichen Sammlungen und Forschungsdaten entwickelt. Konkret handelt es sich dabei um die DARIAH-DE Collection Registry⁹ und das DARIAH-DE Repository sowie um die entsprechenden Datenmodellen, Metadatenstandards und Schnittstellen. Während die Collection Registry den Bereich der wissenschaftlichen Sammlungen abdeckt, worunter neben ausgewählten Beständen traditioneller Gedächtnisorganisationen wie Bibliotheken, Museen und Archiven aber auch Sammlungen geistes- und kulturwissenschaftlicher Forschungsdaten zählen, soll das DARIAH-Repository als Service zur dauerhaften Sicherung, aber auch zur Veröffentlichung von geistes- und kulturwissenschaftlichen Forschungsdaten etabliert werden.

1 <https://de.dariah.eu/steckbrief-und-vision>

2 www.hab.de

3 <https://de.dariah.eu/partner>

4 <https://dev2.dariah.eu/wiki/x/34oYAg>

5 <http://www.sub.uni-goettingen.de/>

6 www.bbaw.de

7 <http://www.dainst.org/>

8 <http://www.ieg-mainz.de/>

9 <https://de.dariah.eu/collection-registry>

Damit die von DARIAH-DE entwickelten Tools und Services aber auch tatsächlich den Bedürfnissen der verschiedenen geistes- und kulturwissenschaftlichen Disziplinen entsprechen können, müssen genau diese Bedürfnisse der Community in Erfahrung gebracht, benannt und soweit abstrahiert und generalisiert werden, dass sich daraus realisierbare Anforderungen an die konkrete Softwareentwicklung ableiten lassen.

Neben der Auswertung bereits existierender Forschungsliteratur und der Ausarbeitung und Durchführung quantitativ orientierter online-Umfragen, stellen ausführlichere Interviews mit ForscherInnen aus den verschiedenen Fachdisziplinen eine weitere Möglichkeiten der Bedarfserhebung für den Aufbau einer digitalen Infrastruktur für Geistes- und Kulturwissenschaften dar, wobei diese drei Zugänge im besten Falle aufeinander abgestimmt sind. Aus diesem Grund werden daher im Anschluss kurz die einschlägige Forschungsliteratur beschrieben und die von DARIAH-DE zum Thema Forschungspraxis und Forschungsdaten bereits durchgeführten Umfragen knapp vorgestellt.

1.2 Forschungsstand

Die Notwendigkeit Forschungsdaten dauerhaft zu speichern und zumindest zum Zwecke der Transparenz der eigenen Forschungsergebnisse zu veröffentlichen, wurde von nationalen wie internationalen Fördergebern mehrmals klar zum Ausdruck gebracht. Exemplarisch dafür sei auf die von der Deutschen Forschungsgemeinschaft 2013 neu aufgelegten „Vorschläge zur Sicherung guter wissenschaftlicher Praxis“ (DFG 2013) und auf die von der Europäischen Kommission veröffentlichten „Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020“ verwiesen (Europäische Kommission 16.12.2013).

Parallel dazu lässt sich auf ein verstärktes Bemühen im Aufbau entsprechender Angebote zur dauerhaften Speicherung und Veröffentlichung von Forschungsdaten feststellen, sei dies auf institutioneller,¹⁰ regionaler, nationaler Ebene,¹¹ egal ob auf einzelne Fächer fokussiert¹² oder offen für Daten unabhängig eines spezifischen disziplinären Erhebungskontextes.¹³ Hand in Hand mit der Konzeption derartiger Datenarchive oder Repositorien erfolgt meist auch eine allgemeintheoretische wie konkret-praktische Auseinandersetzung mit dem Thema Forschungsdaten. Diese Beschäftigung schlägt sich in der Veröffentlichung wissenschaftlicher Artikel und/oder Projektberichten (Wilson 2014, Kindling, Schirnbacher, and Simukovic 2013) ,detaillierten Empfehlungen und Guidelines (League of European Research University 2014; Jones, Pryor and Whyte 2013) bis hin zu kompakten und an ein breites Publikum adressierten Infobroschüren nieder.¹⁴ Einen guten Einstieg in das Thema Archivierung und Publikation von Forschungsdaten stellt die von Merial Patrick zusammengestellte „Data Management Bibliography“ dar, auch wenn die Bibliographie nur Arbeiten bis einschließlich 2010 umfasst (Merial 2010). Dies zeigt jedoch, dass das Thema Forschungsdaten und -speicherung zumindest jenseits des deutschsprachigen Raumes schon seit geraumer Zeit intensiv diskutiert wurde.

Neben technischen und rechtlichen Aspekten, die Stichwörter dazu lauten Langzeitarchivierung und Lizenzen, spielt in den verschiedenen Publikationen häufig auch die Frage nach dem

10 Vgl. dazu etwa HRK 2014. Zum Aufbau eines 'Data Center for the Humanities' an der Universität Köln siehe: Sahle und Kronenwett 2013.

11 <http://www.e-infrastructures.at/das-projekt/>; <http://researchdata.ands.org.au/>; <http://www.re3data.org/> oder auch <http://humanities-data-centre.org/>.

12 <https://datorium.gesis.org/xmlui/>

13 <http://datahub.io/de/>

14 <http://auffinden-zitieren-dokumentieren.de/>

konkreten Bedarf nach Repositorien bzw. nach dem derzeitigen Status-Quo, dem praktischen Umgang mit Forschungsdaten in den verschiedenen Wissenschaftsdisziplinen eine große Rolle. Eine solche Bedarfserhebung kann anhand der Auswertung vorhandener Forschungsliteratur erfolgen. Diesen Weg beschritt etwa Martin Feijen, der 2011 ein Paper mit dem bezeichnenden Titel „What researchers want. A literature study of researchers’ requirements with respect to storage and access to research data“ veröffentlichte (Feijen 2011). Darin werden 14 Umfragen zum Themenkomplex Forschungsdaten und -praxis ausgewertet, die in den Jahren 2008-2010 in den Niederlanden, den USA, England, Australien, Finnland bzw. in ganz Europa durchgeführt worden waren (Feijen 2011, S. 7). Abgesehen von Beobachtungen eher genereller Natur, etwa dass die Notwendigkeit zur Schaffung von Datenrepositorien unbestritten ist, dass diese Repositorien sowohl sicher als auch unkompliziert in der Handhabung sein müssen oder dass die Bedürfnisse und Anforderungen an ein solches Repository von Fachdisziplin zu Fachdisziplin unterschiedlich ausfallen, finden sich in dieser Auswertung auch einige interessante konkrete Beobachtungen.

Die wohl wichtigste Aussage des gesamten Papers, vorausgesetzt man möchte wirklich Services anbieten, die bestmöglich an den tatsächlichen Arbeitsbedingungen der Forscherinnen und Forscher anknüpfen, ist ein Zitat aus dem Report zu der 2009 durchgeführten Umfrage PARSE. Zum Punkt „Data types and storage“ heißt es: „Not surprisingly, MS Office documents are most often used by the respondents. What is a bit surprising perhaps is that still 6% of the respondents do not use MS Office documents.“ (Kuipers and van der Hoeven 2009).

Beispiele von Umfragen und Interviews aktuelleren Datums als jene, die von Feijen ausgewertet wurden, können im Umfeld der Humboldt Universität Berlin sowie im DARIAH-DE Kontext gefunden werden. Diese repräsentieren auch die beiden weiteren Möglichkeiten, Informationen über aktuelle Bedürfnisse von Forscherinnen und Forschern und zu deren Forschungsalltag in Erfahrung zu bringen. In Vorbereitung zum Aufbau eines Datenrepositoriums an der Humboldt Universität wurde der Status Quo im Umgang mit Forschungsdaten sowie der potentielle Bedarf an einem derartigen Repository in Form einer Umfrage abgefragt und die Ergebnisse 2013 veröffentlicht (Simukovic, Kindling, and Schirnbacher 2013). Einzelne Fragen dieser Umfragen wurden in die an der Herzog August Bibliothek durchgeführten Interviews übernommen. Dies trifft zum Teil auch auf Kategorien zur Strukturierung einzelner Antwortmöglichkeiten zu.¹⁵ So erlaubt die konkrete Bezugnahme auf die Umfrage der Humboldt Uni einen präzisen Vergleich der Befragungsergebnisse.

Zusätzlich zu besagter Umfrage, an der 499 Mitarbeiterinnen und Mitarbeiter der Humboldt Uni teilgenommen hatten, wurden zwischen Juli 2013 und Juli 2014 auch noch 17 Interviews zum Thema Forschungsdaten und Forschungsdatenrepositorium durchgeführt (Simukovic et al. 2014). Da der entsprechende Bericht erst nach der Ausarbeitung des Fragenkatalogs zur Interviewserie an der Herzog August Bibliothek veröffentlicht wurde, konnte bei den bis zum jetzigen Zeitpunkt (Jänner 2015) bereits durchgeführten Interviews darauf nicht Bezug genommen werden.

Das „Nutzungsverhalten in den Digital Humanities“ stand bei einer gleichlautenden, von DARIAH-DE im Herbst 2014 durchgeführten Umfrage im Fokus (Stiller, Thoden und Leganovic 2015). Konkret abgefragt wurden dabei, welche Phasen eines möglichst generisch beschriebenen Forschungsprozesses für die einzelne Wissenschaftlerin, für den einzelnen Wissenschaftler relevant sind und welche Software, welche Programme und Tools in den jeweiligen Phasen verwendet werden. Von den Verantwortlichen der Studie wurden dabei, unter Berücksichtigung

¹⁵ Die einzelnen Übernahmen sind im Fragenkatalog gekennzeichnet.

vorhandener Forschungsliteratur (Stiller, Thoden und Leganovic 2015) die Phasen (1.) Planung und Konzeption, (2.) Recherche, (3.) Überarbeitung und Schärfung der Forschungsfrage und (4.) Analyse, Annotation und Interpretation identifiziert. Die Idee, den Forschungsprozess in einzelne, möglichst klar voneinander abgrenzbare Arbeitsschritte zu unterteilen, wurde von dieser Umfrage aufgegriffen und für den Fragenkatalog der Interviewreihe weiter ausgebaut.

1.3 Besonderheiten dieser Interviewreihe

Die Besonderheit der hier vorgestellten Interviewreihe liegt darin, dass hier fast ausschließlich ForscherInnen aus den Geistes- und Kulturwissenschaften befragt wurden. Dies ist insofern von Relevanz, als DARIAH-DE sich bekanntermaßen dem Aufbau von Infrastruktur für digitale Geistes- und Kulturwissenschaften verschrieben hat. Für eine Community also, die, wie noch zu zeigen sein wird, über keine sehr klaren Vorstellungen zum Konzept Forschungsdaten verfügt

Ausschließlich Geistes- und KulturwissenschaftlerInnen waren hingegen die Zielgruppe der DARIAH-DE Umfrage „Nutzungsverhalten in den Digital Humanities“. Dabei handelt es sich allerdings um eine mittels Web-Formular durchgeführte online-Umfrage. Eine solche online-Umfrage hat den großen Vorteil, dass dadurch ungleich mehr Personen erreicht werden, als bei einer ortsgebundenen Interviewreihe wie der hier vorgestellten. Mit diesem quantitativen Mehrwert sind jedoch meist auch qualitative Einbußen verbunden. Erstens erlaubt eine offene Gesprächssituation ggf. gezieltes Nachfragen bzw. die spontane Umformulierung von Fragen, weshalb reflektiertere Antworten zu erwarten sind, als dies bei einer online-Umfrage der Fall sein dürfte. Und zweitens liegt die Last der Dokumentation der Antworten nicht bei den Befragten, die sich somit gänzlich auf die Inhalte ihrer Antworten konzentrieren können und nicht etwa überlegen müssen, wie ihre Gedanken und Überlegungen nun am besten in einem Web-Formular unterzubringen sind.

Diese beiden Vermutungen wurden im Zuge der Interviews bestätigt. Dabei muss an dieser Stelle aber darauf hingewiesen werden, dass die an der Herzog August Bibliothek vorherrschenden Arbeitsbedingungen sich auf die Durchführung einer Interviewserie außerordentlich positiv auswirken. StipendiatInnen und GastwissenschaftlerInnen, die nach Wolfenbüttel kommen, um mit den Beständen der Herzog August Bibliothek zu arbeiten, verbringen die meiste Zeit ihres Aufenthaltes im intensiven Studium, weshalb die jeweiligen Interviews – so zumindest der entstandene Eindruck – als äußerst willkommene Ablenkung empfunden wurde. Dies schlug sich auch in der Dauer der Interviews nieder. Die meisten Gespräche dauerten zumeist um die 45 Minuten, konnten sich gelegentlich aber auch auf über 90 Minuten erstrecken.

1.4 Technische Umsetzung

Zur Strukturierung der Interviews und damit einhergehend auch zur Dokumentation der Antworten wurde ein detaillierter Fragenkatalog ausformuliert. Aus diesem Katalog wurde anschließend unter Verwendung des XML-Standards für Formulare XForms¹⁶ und unter Einbindung von XSLTForms¹⁷ eine Eingabemaske für die gängigen Browser generiert. Die einzelnen Antworten können so unmittelbar während der Interviews vom Interviewer mit Hilfe dieser Maske protokolliert werden. Dies macht aufwendiges Mitschreiben oder das Aufnehmen der Interviews auf Tonband samt anschließender Transkription der Aufnahmen überflüssig. Gleichzeitig sorgt das Formular dafür, dass die Antworten stets in der gleichen Reihenfolge und in

¹⁶ <http://www.w3.org/MarkUp/Forms/>

¹⁷ <http://en.wikibooks.org/wiki/XSLTForms/Introduction>

weitgehend gleicher Form festgehalten werden,¹⁸ wodurch auch gleich die Basis für eine automatisierte Auswertung der Interviews errichtet wird.

Nach Abschluss des Interviews generiert das XForms-Dokument eine XML-Datei, die in einer eXist-db XML-Datenbank¹⁹ gespeichert wird, wobei diese eXist-db XML-Datenbank gleichzeitig auch die Generierung der Eingabemaske sowie generell die gesamte Datenverwaltung, -analyse und -anzeige übernimmt. Die erstellten und gespeicherten XML-Dateien der einzelnen Interviews werden dann mittels xQuery analysiert und die Ergebnisse als HTML im Browser angezeigt. Die Analyse und Auswertung der Interviews bzw. die Präsentation der Ergebnisse ist somit 'on the fly' möglich und kann bequem online abgerufen und geteilt werden. Eine gemeinsame Veröffentlichung der Skripte und der Daten gewährleistet außerdem größtmögliche Transparenz und Nachvollziehbarkeit der Analyseergebnisse.

Dieses System wurde ‚quick and dirty‘ entwickelt, erwies sich jedoch als äußerst praktikabel. Im Falle einer weiteren Verwendung sollte das System aber an einigen Stellen überarbeitet werden. Abgesehen von kleineren inhaltlichen Korrekturen wären vor allem Auswahllisten zur Unterstützung einer kontrollierten Dateneingabe zu implementieren und auch die Präsentation der Ergebnisse ist noch ausbaufähig.

2 Ergebnisse I: Programme, Dateiformate & Arbeitsschritte

Eine detaillierte Analyse der Interviews auf Basis der jeweils aktuellen Anzahl durchgeführter Befragungen kann online unter folgendem Link abgerufen werden:

<http://digital-archiv.at:8081/exist/apps/DARIAH-Collection/pages/InterviewAuswertung.html>

In diesem Paper werden daher nur die aus diesem Material erkennbaren großen Trends beschrieben.

2.1 MS-Word und/oder Citavi, Zotero & Co

Tendenziell können die befragten Geistes- und Kulturwissenschaftlerinnen in zwei Gruppen unterteilt werden. Während die eine Gruppe für die Arbeitsschritte Bibliographieren, Exzerpieren und Notieren fast ausschließlich auf MS-Word setzt,²⁰ verwendet die andere Gruppe für diese Tätigkeiten eines von mehreren Literaturverwaltungsprogrammen, am häufigsten Zotero und Citavi, genannt wurden aber auch Endnote, OneNote, Filemaker und Litlink. Zumindest zum Schreiben wird aber auch in der zweiten Gruppe beinahe ausschließlich MS-Word verwendet, wobei dafür Texte oder bibliographische Angaben wie Voll- oder Kurzzitate aus den Literaturverwaltungsprogrammen nach MS-Word importiert werden.²¹

18 Dass die in die Datenmaske eingetragenen Werte eine einheitliche Form aufweisen, liegt vorerst noch in der Verantwortung und Disziplin des Interviewers. Da in den meisten Fällen keine Auswahl aus vorgefertigten Antworten möglich ist, obliegt es nämlich dem Interviewer, ob er nun etwa „MS-Word“, „MS Word“ oder vielleicht nur „Word“ schreibt. Aufgrund der in den bisherigen Interviews angesammelten Antwortmöglichkeiten wäre jedoch zu überlegen, ob diese nicht als normalisierte Antwortmöglichkeiten in die Eingabemaske implementiert werden sollten.

19 <http://exist-db.org/>

20 Eine Befragte gab an, nicht MS Word sondern OpenOffice zu verwenden.

21 Dieser Befund wird von den Ergebnissen der DARIAH-Umfrage klar bestätigt. Vgl. Stiller et al. 2015, S. 28f.

Bei dem am häufigsten verwendeten Dateiformat handelt es sich daher konsequenterweise um doc bzw. docx, wobei einzelne Forscherinnen und Forscher aber auch angaben, gelegentlich andere Dateiformate zu verwenden, um möglichen Kompatibilitätsproblemen aus dem Weg zu gehen. In diesem Kontext genannt wurden rtf und odt genannt. Da sowohl MS-Word als auch die genannten Literaturverwaltungsprogramme in der Lage sind, die von ihnen erstellten/verwalteten Dateien und Datensätze in mehrere unterschiedliche Dateiformate zu exportieren, ist eine Frage nach den verwendeten Dateiformaten indirekt meist eine Frage nach den verwendeten Programmen bzw. kann anhand der genannten Programme auch auf die potentiell anzutreffenden Dateiformate geschlossen werden.

2.2 Datenmanagement: Windows/Mac-Explorer oder Citavi, Zotero & Co

Wo die zweite Gruppe die Ordnung und Strukturierung der im Laufe des Forschungsprozesses entstehenden Bibliographien, Notizen und Exzerpte dem Literaturverwaltungsprogramm ihrer Wahl überlassen, entwickelt die erste Gruppe ihr eigenes Ordnungssystem durch Anlegen einer hierarchischen Ordnerstruktur im Windows-Explorer bzw. dem Mac-OS Pendant.

Als zentrales Element dieses Ordnungssystems, als atomare Einheit, sei dies nun ein Datensatz in einem Literaturverwaltungsprogramm oder eine im Windows-Explorer gespeicherte Datei, kann in den meisten Fällen ein konkretes Objekt, ein Buch, ein Kupferstich, ein Archivadokument identifiziert werden. Weniger häufig konnte hingegen beobachtet werden, dass einzelne Themen oder Schlagwörter solche atomaren Einheiten bilden.

Dies ergab sich aus einzelnen konkreten Zusatzfragen, die nicht im Fragenkatalog vorgesehen waren, weshalb zu diesem Punkt keine überprüfbare Dokumentation vorliegt. Hierin zeigt sich der Nachteil des Verzichtes auf eine vollständige Aufzeichnung der einzelnen Interviews. Allerdings kann die Frage nach der Ordnungsstruktur der eigenen Dateien bzw. Datensätze in eine überarbeitete Version des Fragenkataloges aufgenommen werden.

2.3 Typische Arbeitsschritte: Bibliographieren, Notieren & Schreiben

Bibliographieren, Exzerpieren/Notieren und Schreiben sind die wichtigsten Arbeitsschritte in den Geistes- und Kulturwissenschaften. Diesen Eindruck vermitteln zumindest die bisher geführten Interviews. Sämtliche der bisher befragten Forscherinnen und Forscher gaben an, diese Tätigkeiten im Zuge ihrer Forschungen auszuüben. Aus diesem Grund erscheint es angebracht, diese Arbeitsschritte als typische Arbeitsschritte innerhalb eines geistes- und kulturwissenschaftlichen Forschungsprozesses zu bezeichnen.

Als wichtiger Arbeitsschritt konnte im Zuge der Gespräche und Diskussionen auch noch die Verwaltung der Forschungsmaterialien identifiziert werden. Wie bereits angeführt, wird dafür entweder ausschließlich auf das Dateimanagementsystem des Betriebssystems zurückgegriffen oder zusätzlich noch ein Literaturverwaltungssystem verwendet.

Eine Diskussion zu diesem Thema ergab sich meist anlässlich der Frage 3.4 „Welche Programme verwenden Sie zum Sammeln und/oder zum Strukturieren von Forschungsdaten?“. Eine Frage, die eigentlich darauf abzielte, in Erfahrung zu bringen, welche Programme zur Strukturierung und Analyse quantitativer Datensätze (Steuerlisten, Statistiken zur Bevölkerungsentwicklung oder Ergebnisse von Textmining) verwendet werden.

Im Falle einer Überarbeitung des Fragebogens erscheint es demnach ratsam, den Punkt Datenverwaltung und -organisation als eigenen Arbeitsschritt zu behandeln, wobei hier auch gezielt nach der vermehrt praktizierten Verwaltungslogik (objekt- oder themenzentriert) gefragt werden könnte.

2.4 Spezifische Arbeitsschritte

Während die typischen Arbeitsschritte in allen Projekten anzutreffenden sind, wurden die übrigen Arbeitsschritte nicht mehr von allen Befragten als relevant eingeschätzt, wobei es hier klare Unterschiede zwischen den verbleibenden Tätigkeiten gibt. Aufgaben wie Transkribieren, „Bearbeiten, Ansehen und Speichern von Bildern und insbesondere Präsentieren spielen in ungleich mehr Projekten eine Rolle als die Analyse quantitativer Datensets, Zeichnen und Skizzieren, Bearbeiten von Audiofiles, oder das Schreiben von Skripten und Computerprogrammen. Was die dabei jeweils verwendeten Programme (und Dateiformate) betrifft, sei hinsichtlich der jeweiligen Daten auf die online Auswertung der Interviews verwiesen.

Kritisch anzumerken ist hierbei jedoch, dass der Fragenkatalog streng genommen keine Fragen nach der Bewertung der genannten Arbeitsschritte enthält, sondern nur welche Programme für die Bewältigung der genannten Aufgaben verwendet werden. Außerdem gilt es zu bedenken, dass die Befragten in Wolfenbüttel meist sehr textlastige Projekte verfolgen. Eine entsprechende Umformulierung des Fragenkataloges sowie eine Ausweitung der Interviews auf andere Institutionen könnten zu aussagekräftigeren Ergebnissen führen.

Trotz dieser methodischen Mängel scheint die hier vorgenommene Trennung zwischen typischen und spezifischen Arbeitsschritten für geistes- und kulturwissenschaftliche Forschungsprojekte plausibel und sollte eingehender verfolgt werden.

3 Ergebnisse II: Forschungsdaten in den Geistes- und Kulturwissenschaften

3.1 Forschungsdaten zum Zwecke der Nachvollziehbarkeit

Im Zuge der Interviews zeigte sich, dass die Befragten Schwierigkeiten mit dem Konzept „Forschungsdaten“ hatten, vor allem hinsichtlich der Abgrenzung von „Quelle“ und „Publikation“ und im Zusammenhang mit der Forderung nach der Publikation von Forschungsdaten, um Forschungsergebnisse nachvollziehbar und überprüfbar zu machen.

Der Hauptgrund dafür muss mit großer Wahrscheinlichkeit in der Arbeitsweise der einzelnen Wissenschaftlerinnen und Wissenschaftler zu suchen sein. Entsprechend ihrer (geistes- und kultur)wissenschaftlichen Sozialisierung steht am Abschluss des Forschungsprojektes die Veröffentlichung eines Textes, einer wissenschaftlichen Publikation. Diese enthält Informationen zur verfolgten Fragestellung, der verwendeten Sekundärliteratur, der untersuchten Quellen samt einer Begründung der Auswahl derselben, sowie eine, im besten Falle kritische und selbstreflexive Schilderung der Arbeitsschritte und Gedankengänge, welche die Nachvollziehbarkeit der daraus gewonnenen und veröffentlichten Erkenntnisse gewährleisten sollte.

Angesichts dieser für die Geistes- und Kulturwissenschaften typischen Arbeitsweise überrascht es wenig, dass eben jene Fragen, in denen es um die Veröffentlichung von Forschungsdaten zum

Zwecke der Nachvollziehbar- und Überprüfbarkeit geht, Schwierigkeiten bereiteten.²² Denn im Selbstverständnis der ForscherInnen ist die Überprüfbarkeit der Ergebnisse durch die Publikation (in der eben beschrieben Form) gewährleistet. Seitens der Interviewten wurde in diesem Kontext mehrmals explizit auf die integralen Bestandteile wissenschaftlicher Publikationen, nämlich Fußnotenapparat, Quellen- und Literaturverzeichnis hingewiesen.

Aus dieser Beobachtung ergeben sich folgende Konsequenzen: Dort wo in den Geistes- und Kulturwissenschaften qualitativ bzw. exemplarisch gearbeitet wird, wo Verstehen und Interpretieren im Zentrum des Erkenntnisinteresses stehen, kurz wo hermeneutische Methoden dominieren, ist alleine aus Gründen der Nachvollziehbarkeit und Überprüfbarkeit kein Bedarf an Forschungsdatenrepositorien erkennbar, da Nachvollziehbarkeit und Überprüfbarkeit durch die gängige Praxis wissenschaftlichen Publizierens in den Geisteswissenschaften gewährleistet ist. Im Gegenzug bedeutet dies aber auch, dass in jenen Bereichen, in denen qualitative bzw. empirische Methoden eingesetzt werden, ein Bedarf an entsprechenden Repositorien gerade auch unter dem Gesichtspunkt der Gewährleistung von Überprüfbarkeit notwendig sind, da dies in Form der traditionellen wissenschaftlichen Publikation nicht oder nur auf wenig zufriedenstellende Art und Weise, etwa in der Form von seitenlangen tabellarischen Anhängen möglich ist.

Bedauerlicherweise kann aufgrund der im Zuge der Interviews bisher gesammelten Daten nicht eindeutig eruiert werden, wie viele der befragten Forscherinnen und Forscher im Zuge ihrer Projekte auf quantitative Methoden setzen, da jene Frage, die darauf abzielen sollte, auch Informationen hinsichtlich der Strukturierung des eigenen Forschungsmaterials einholte.

3.2 Forschungsdaten zum Zwecke der Nachnutzung

Der Begriff Forschungsdaten spielt bei einer Betonung des Aspekts Überprüfbarkeit und Nachvollziehbarkeit von Forschungsergebnissen im Kontext traditioneller, sprich überwiegend hermeneutisch arbeitender Geistes- und Kulturwissenschaften so gut wie keine Rolle. Rücken jedoch die Aspekte kollaboratives Arbeiten und Nachnutzbarkeit von Forschungsdaten in den Vordergrund, ändert sich dieses Bild.

So bejahten neun der bisher 15 Interviewten die Frage: „Sollen Ihre Forschungsdaten von anderen Personen verwendet werden dürfen?“ und drei weitere gaben an, dies von den Daten abhängig machen zu wollen.²³ Kategorisch schlossen nur drei Personen die Nachnutzung ihrer Daten aus. Unter Forschungsdaten wurden dabei all jene Daten bzw. Dateien verstanden, die im Zuge des Forschungsprozesses gesammelt und erstellt wurden und die nicht in weitgehend unveränderter Form Teil der wissenschaftlichen Publikation sind. In den meisten Fällen handelte es sich dabei um Exzerpte von Quellen und Sekundärliteratur, Transkripte von Quellen, Notizen und Anmerkungen zu konkreten Texten/Bildern/Objekten oder zu Ideen, Thesen, Konzepten, (kommentierte und/oder strukturierte) Bibliographien sowie (mehr oder weniger legale) digitale Reproduktionen von Fachliteratur und/oder Quellen, ggf. mit Anmerkungen versehen.

22 Konkret handelt es sich dabei um die Fragen III. 15: „Welche Forschungsdaten würden Sie dort [in einem Datenarchiv, Anm.] ablegen? Bitte wählen Sie alle zutreffenden Antworten aus“, wobei diese Frage aus der an der Humboldt-Universität durchgeführten Umfrage übernommen wurde. Vgl. Simukovic, Kindling, and Schirnbacher. 2015, S. 54.

23 Diese Antwortoption wurde erst nach der Durchführung der ersten rund sieben bis acht Interviews in den Fragenkatalog aufgenommen.

Die Hauptgründe weshalb sich einige der Interviewten gegen eine Weitergabe einiger oder gar aller Daten aussprachen, liegen einerseits in der als zu gering erachteten Qualität der Daten und andererseits in zukünftigen Projekten/Publicationen, die auf diesen Daten basieren sollen.

Sollten Daten zum Zwecke der Nachnutzung veröffentlicht werden, dann erwarteten elf der 15 Interviewten, in den entsprechenden Publikationen zitiert zu werden und vier Personen wünschen sich außerdem noch darüber informiert zu werden, wenn andere die von ihnen bereitgestellten Daten ansehen oder herunterladen.²⁴

Aufschlussreich hinsichtlich der Schärfung des Begriffes Forschungsdaten erwies sich letztlich auch die Frage, ob man selbst schon einmal mit „nicht publizierten, individuell erhobenen Forschungsdaten anderer“ gearbeitet habe, wobei diese Formulierung „nicht publiziert...“ selbst ein Ergebnis von Diskussionen ist, die während der ersten Interviews im Kontext dieser Frage geführt wurden, wo nur nach der Verwendung von Forschungsdaten anderer gefragt worden war, was oftmals die Gegenfrage evozierte, was genau nun darunter zu verstehen sei. In einer ersten Reaktion war die Fragen nach der Verwendung von Forschungsdaten anderer nämlich meist sehr entschieden und ohne langes Nachdenken abgelehnt worden und erst die präzisere Formulierung, die in den ersten Interviews noch mündlich erfolgte und schließlich in der oben zitierten Form in den Fragebogen einfluss, führte zu dem vorliegenden Ergebnis demzufolge elf der fünfzehn Befragten angaben, schon einmal mit Forschungsdaten anderer gearbeitet zu haben. Am häufigsten handelte es sich dabei um Bibliographien bzw. im direkten Gespräch vermittelte Literaturhinweise, sowie um Bilddateien von Quellen und Sekundärliteratur. Genannt wurden außerdem noch Exzerpte, Transkripte und statistisches Material. Dieser Datenaustausch findet dabei jedoch stets auf persönlicher Ebene statt. Dieser im Zuge der Diskussion gewonnene Erkenntniswert wird vom Fragebogen zwar nicht explizit dokumentiert, allerdings machen die Antworten auf die Frage nach der Verwendung von institutionellen und öffentlichen Datenrepositorien deutlich,²⁵ dass der eben beschriebene Datenaustausch zwangsläufig über persönliche Netzwerke und Bekanntschaften funktionieren muss. So gab keiner der Befragten an, derartige Dienste in Anspruch zu nehmen. Gleichzeitig sagten aber zehn Forscherinnen und Forscher, zumindest manche ihrer Daten in einem derartigen Repository ablegen zu wollen.

4 Abschluss und Ausblick

Die hier vorgenommene Evaluierung unter den (Gast)Wissenschaftlern an der Herzog August Bibliothek zeigt, dass in tendenziell traditionellen, textbasierten und vornehmlich exemplarisch, verstehend, hermeneutisch arbeitenden Geistes- und Kulturwissenschaften, Bedarf an Forschungsdatenrepositorien mehr zum Zwecke der Förderung kollaborativen Arbeitens und zur Generierung von Möglichkeiten zur Nachnutzung von Forschungsdaten die von Einzelforschern erhoben wurden besteht. So lange empirisch-quantitative Methoden aber nur wenig im Fokus Geistes- und kulturwissenschaftlichen Arbeitens stehen – wie dies bei den auf den Beständen der

24 Dass vier Personen an eine Veröffentlichung ihrer Daten keinerlei Bedingungen geknüpft haben, wie es die Auswertung der Interviews nahelegt, ist auf eine leider gegebene Inkonsequenz der Dateneingabemaske zurückzuführen. Die korrekte Anzahl ergibt sich hier vielmehr aus der Differenz der Anzahl jener Personen, die sich eine Veröffentlichung von Daten vorstellen können und der Summe der Antworten hinsichtlich der an eine Veröffentlichung geknüpften Bedingungen. Somit ergibt sich, dass sich nur eine Person eine bedingungslose Veröffentlichung der eigenen Forschungsdaten vorstellen kann.

25 Frage V. 8: „Verwenden Sie zur Datensicherung/Publication spezielle (fachspezifische) Clouddienste und Datenarchive?“

Herzog August Bibliothek basierenden Forschungen überwiegend der Fall ist, ist der Bedarf an einem Repositorium zur reinen Dokumentation der Forschungsergebnisse nicht gegeben.

Was die Art, den Typ der produzierten Forschungsdaten anbelangt, so handelt es sich dabei fast zur Gänze um Textdateien. Diese liegen dabei in der Regel als mit MS-Word erstellte docx-Dateien oder als Datensätze in einem Literaturverwaltungsprogramm vor. Wo kein Literaturverwaltungsprogramm (Citavi, Zotero...) zum Einsatz kommt, wird die Ordnung und Strukturierung der Dateien in der Regel vom Dateiverwaltungssystem des Browsers übernommen. Fragestellungen, Materialien und Methoden mögen innerhalb der Geistes- und Kulturwissenschaften zwar gerne über die so häufig postulierte Vielfalt aufweisen, lassen sich aber doch auf einen breite und einheitliche Basis herunterbrechen. Eine Basis, auf der gleichzeitig auch die größte Menge an Forschungsdaten anfällt. Forschungsdaten, die mit sehr ähnlichen Programmen und auf sehr ähnliche Art und Weise erzeugt, organisiert und gespeichert werden, weshalb eine Überführung dieser Daten in ein öffentlich zugängliches, vertrauenswürdiges und dauerhaftes Systems keine allzu großen technischen Hürden zu überwinden haben dürfte.

Dieser Befund beruht auf der überschaubaren Menge von derzeit 15 Interviews, die noch dazu an einer einzigen Forschungseinrichtung geführt wurden, deren materieller Schwerpunkt auf der Quellengattung Buch liegt und die auf die frühe Neuzeit spezialisiert ist. Gleichzeitig wurde für die Durchführung und Dokumentation der Interviews aber ein leicht zu handhabender Workflow entwickelt (Fragenbogen + Eingabemaske/Dokumentation + automatisierte Auswertung). Dieser kann nach einer Einarbeitung der in diesem Bericht angemerkten Optimierungsvorschläge ohne großen Aufwand von anderen Forschungseinrichtungen – zu denken wäre hier in erster Linie an die Projektpartner innerhalb DARIAH-DEs wie das IEG-Mainz, das Deutsche Archäologische Institut, das Musikwissenschaftliche Seminar Detmold/Paderborn sowie die eingebunden Universitäten – übernommen werden. Denn dann wäre es möglich, einen tieferen, präziseren und vor allem auch repräsentativeren Einblick zum Thema „Forschen und Forschungsdaten in den Geistes- und Kulturwissenschaften“ bekommen zu können, als dies anhand singulärer Interviewreihen oder großer aber sehr weit gefasster Onlineumfragen möglich ist.

5 Literaturverzeichnis

DFG. *Vorschläge zur Sicherung guter wissenschaftlicher Praxis*. Ergänzte Auflage. Bonn, 2013.

http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf.

Europäische Kommission. *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020. Version 16 December 2013*, 2013.

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.

Feijen, Martin. *What Researchers Want. A Literature Study of Researchers' Requirements with Respect to Storage and Access to Research Data.*, 2011.

HRK Hochschulrektorenkonferenz. *Management von Forschungsdaten - Eine zentrale strategische Herausforderung für Hochschulleitungen. Empfehlungen der 16. Mitgliederversammlung der HRK Am 13. Mai 2014 in Frankfurt am Main*. Frankfurt am Main, 2014.

http://www.hrk.de/uploads/tx_szconvention/HRK_Empfehlung_Forschungsdaten_13052014_01.pdf.

Jones, Sarah, Graham Pryor, and Angus Whyte. *How to Develop Research Data Management*

- Services – a Guide for HEIs*. Edinburgh, 2013. <http://www.dcc.ac.uk/resources/how-guides/how-develop-rdm-services>.
- Kindling, Maxi, Peter Schirmbacher, and Elena Simukovic. "Forschungsdatenmanagement an Hochschulen: Das Beispiel der Humboldt-Universität zu Berlin." *Libreas. Library Ideas* 23 (2013). <http://edoc.hu-berlin.de/libreas/23/kindling-maxi-1/PDF/kindling.pdf>.
- Kuipers, Tom, and Jeffrey van der Hoeven. *Insight into Digital Preservation of Research Output in Europe (PARSE). Survey Report*, 2009. http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf.
- League of European Research University. *LERU Roadmap for Research Data*, 2014. http://www.uzh.ch/research/LERU_Roadmap_for_Research_data.pdf.
- Merial, Patrick. *Data Management Bibliography*. Oxford University, 2010. <http://sudamih.oucs.ox.ac.uk/docs/Data%20management%20bibliography.pdf>.
- Sahle, Patrick, and Simone Kronenwett. "Jenseits der Daten. Überlegungen zu Datenzentren für die Geisteswissenschaften am Beispiel des Kölner ‚Data Center for the Humanities‘." *Libreas* 23 (2013): 76–96.
- Simukovic, Elena, Maxi Kindling, and Peter Schirmbacher. *Forschungsdaten an der Humboldt Universität zu Berlin. Bericht über die Ergebnisse der Umfrage zum Umgang mit digitalen Forschungsdaten an der Humboldt-Universität zu Berlin*, 2013.
- Simukovic, Elena, Raphael Thiele, Alexander Struck, Maxi Kindling, and Peter Schirmbacher. *Was Sind Ihre Forschungsdaten? Interviews mit Wissenschaftlern der Humboldt-Universität zu Berlin*. Berlin, 2014. <http://edoc.hu-berlin.de/oa/reports/rem4TV5CmzMjE/PDF/27ql5s1RXm17k.pdf>.
- Stiller, Juliane, Klaus Thoden, Oona Leganovic, Christian Heise, Mareike Höckendorff, and Timo Gnad. *Nutzungsverhalten in den Digital Humanities*, 2015. <https://dev2.dariah.eu/wiki/download/attachments/26150061/Report1.2.1-final.pdf?version=6&modificationDate=1430303390972&api=v2>.
- Sturges, Paul. "Zugang zu Forschungsdaten. Die Notwendigkeit von Richtlinien für Data Sharing bei Fachzeitschriften. Das JoRD-Projekt der Universität Nottingham." *B.I.T. online: Zeitschrift für Bibliothek, Information und Technologie* 5, no. 17/2014 (2014): 421–30.
- Wilson, James A. J. *JISC Final Report, Supporting Data Management Infrastructure for the Humanities*. Oxford University, 2011. http://sudamih.oucs.ox.ac.uk/docs/Sudamih_FinalReport_v1.0.pdf.

6 Links

<http://auffinden-zitieren-dokumentieren.de/>

<http://datahub.io/de/>

<http://digital-archiv.at:8081/exist/apps/DARIAH-Collection/pages/InterviewAuswertung.html>

<http://en.wikibooks.org/wiki/XSLTForms/Introduction>

<http://exist-db.org/>

<http://hab.de/>

<http://humanities-data-centre.org/>

<http://researchdata.ands.org.au/>

<http://www.bbaw.de/>

<http://www.dainst.org/>

<http://www.e-infrastructures.at/das-projekt/>

<http://www.ieg-mainz.de/>

<http://www.re3data.org/>

<http://www.sub.uni-goettingen.de/>

<http://www.w3.org/MarkUp/Forms/>

<https://datorium.gesis.org/xmlui/>

<https://de.dariah.eu/collection-registry>

<https://de.dariah.eu/partner>

<https://de.dariah.eu/steckbrief-und-vision>

<https://dev2.dariah.eu/wiki/x/34oYAg>

Alle Links aufgerufen am 8. Mai 2015.