

Forschungsdaten in den Geisteswissenschaften

Bereits selbstverständlich oder doch noch etwas exotisch?

Susanne Blumesberger, Universität Wien

Zusammenfassung

Der Beitrag konzentriert sich bewusst auf jene Daten in den Geisteswissenschaften, die sich nicht den Digital Humanities zuordnen lassen und damit nicht schon per se bewusst auf die digitale Dimension des Fachs fokussieren. Spätestens beim Ausfüllen der von den Fördergebern vorgeschriebenen Datenmanagementplänen stellt sich für viele Wissenschaftler*innen die Frage, was denn nun in ihrem Fall eigentlich die Forschungsdaten sind, die sie sicher und langfristig verfügbar machen sollen. Handelt es sich dabei um die Texte die verwendet oder produziert werden oder die Bilder, auf die der Text Bezug nimmt? Die fehlende Auseinandersetzung mit Datenmanagement während des Studiums und beim Verfassen der Abschlussarbeit führt nicht nur dazu, dass viele Informationen, die von anderen Forscher*innen weitergenutzt werden könnten, verloren gehen, sondern auch zu Rechtsunsicherheiten und im schlimmsten Fall zu juristischen Problemen. Hier den richtigen Weg zwischen aus Furcht vor juristischen Konsequenzen zurückgehaltenen Informationen und einem möglichst offenen Umgang mit Daten zu finden, ist eine Gratwanderung und bedarf genauer Anleitungen und viel Hintergrundwissen. Wie man als Forschungsdatenmanager*in an Bibliotheken am besten Geisteswissenschaftler*innen dabei hilft „as open as possible, as closed as necessary“ zu agieren, soll im Rahmen des Beitrags andiskutiert werden.

Summary

This article deliberately focuses on data in the humanities that cannot be classified as digital humanities and thus does not per se consciously focus on the digital dimension of the discipline. At the latest when filling out the data management plans prescribed by the funding bodies, many scientists and scholars are confronted with the question of what actually constitutes the research data that they are supposed to make available securely and in the long term. Is it the text that is used or produced, or the images that the text refers to? Not dealing with data management while studying and writing the thesis not only leads to the loss of a lot of information that could be further used by other researchers, but also to legal uncertainties and, in the worst case, to legal problems. Finding the right path between withholding information for fear of legal consequences and handling data as openly as possible is a tightrope walk and requires precise instructions and a lot of background knowledge. This article will discuss how research data managers in libraries can best help humanities scholars to act “as open as possible, as closed as necessary”.

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/5739>

Blumesberger, Susanne: ORCID: <https://orcid.org/0000-0001-9018-623X>

Schlagwörter: Forschungsdaten; Geisteswissenschaften; Forschungsdatenmanagement; Humanities; Universitätsbibliothek Wien

Dieses Werk steht unter der Lizenz [Creative Commons Namensnennung 4.0 International](https://creativecommons.org/licenses/by/4.0/).

Was sind meine Forschungsdaten?

Sind Geisteswissenschaftler*innen mit einem Datenmanagementplan¹, der vom größten österreichischen Fördergeber, dem FWF, seit 1.1.2019 nach der Zusage eines Projekts gefordert wird, konfrontiert, bzw. wird eine Forschungsdatenpolicy in der eigenen Institution veröffentlicht, stellen sich viele von ihnen Fragen, die sie bisher nicht mitgedacht haben, weil sie ihnen bisher nicht relevant genug erschienen. Eine der Fragen ist beispielsweise was denn nun die Forschungsdaten sind, die sie nach den FAIR-Prinzipien² – auch diese sind oft unbekannt – behandeln sollen. Handelt es sich dabei um die Texte die verwendet oder produziert werden oder um die Bilder, auf die der Text Bezug nimmt, oder was ist damit genau gemeint? In vertiefenden Gesprächen mit Forschenden wird deutlich, dass der Begriff „Forschungsdaten“ kaum von den Wissenschaftler*innen aus den Humanities verwendet wird. Was beispielsweise für Natur- oder Sozialwissenschaftler*innen zum täglichen Sprachgebrauch gehört, ist bei Geisteswissenschaftler*innen eher fremd, denn Bilder, Texte, Aufzeichnungen, werden, auch wenn sie digital vorhanden sind, kaum als Daten wahrgenommen. Unklar ist auch, welche Arten von Daten in den Datenmanagementplänen, bzw. Policies gemeint sind, denn viele Digitalisate dienen ja ausschließlich der eigenen Verwendung, wie beispielsweise Abbildungen aus Büchern oder gescannte Archivmaterialien. Folgt man der Definition des österreichischen Wissenschaftsfonds, sind im Datenmanagementplan vor allem jene Daten angesprochen, die einer Publikation zugrunde liegen. Diese müssen, wenn möglich frei verfügbar gemacht werden, unter anderem auch mit dem Ziel den Forschungsprozess transparent und die Ergebnisse gegebenenfalls überprüfbar zu machen. Aber auch alle anderen Daten, die für die Forschung verwendet werden, müssen in einem Datenmanagementplan beschrieben werden, auch wenn sie aus unterschiedlichen Gründen nicht frei verfügbar gemacht werden können, weil sie beispielsweise nur als Arbeitsgrundlage zur Verfügung gestellt, jedoch nicht für die Veröffentlichung freigegeben wurden. Bei diesen Daten ist anzugeben, was gegen eine Veröffentlichung spricht, das können juristische Gründe sein, weil beispielsweise die Schutzfrist für urheberrechtlich geschützte Werke noch nicht abgelaufen ist. Es können aber auch ethische Gründe dagegensprechen, wenn zum Beispiel in einem Nachlass intime Details über Dritte preisgegeben werden, die man nicht verbreiten möchte. In den meisten Fällen klärt sich bei beratenden Gesprächen recht rasch auf, dass unter dem oft als sehr technisch empfundenen Begriff „Daten“ vielleicht auch nur die Arbeitsgrundlagen verstanden werden, die Basis der täglichen Arbeit sind.

Was bedeutet FAIR?

Auch die Frage des FWF „Welche Informationen werden benötigt, damit die Daten in Zukunft auffindbar, zugänglich, interoperabel und wiederverwendbar (FAIR) sind?“ löst bei vielen Forschenden in den Geisteswissenschaften zunächst Unbehagen aus. Hier werden Begriffe verwendet, die bisher im Alltag der Wissenschaftler*innen kaum eine Rolle gespielt haben. Bis vor wenigen Jahren stand vor allem in den Geisteswissenschaften am Ende eines Forschungsprozesses eine Publikation, meist

1 Siehe auch: Blumesberger, Susanne: The data management plan - the common thread through the project. DMP consulting at the University of Vienna. In: Proceedings of Science, Austrian Citizen Science Conference 2020 – ACSC2020, 2021.

2 <<https://www.go-fair.org/fair-principles/>> (5.7.2021).

in Form einer Monographie oder als Beitrag in einem Sammelband, bzw. einer wissenschaftlichen Zeitschrift. Der Endbericht für den Fördergeber belegte zwar den Prozess dorthin, beschrieb die Methoden und die Quellen, aber nach den zugrundeliegenden Materialien fragte nach Projektende im Regelfall niemand mehr. Aufzeichnungen, Kopien, Scans, Bilder, Transkripte, Audio- oder Videoaufnahmen wurden entweder entsorgt oder in privaten oder dienstlichen Arbeitszimmern aufbewahrt bis sie schließlich in den Nachlass von prominenten Wissenschaftler*innen übergingen und bei prominenteren Wissenschaftler*innen im Nachlass landeten, der zum Teil in Bibliotheken und Archiven aufbewahrt wurde und jetzt zum Teil nach und nach aufbereitet und zur Verfügung gestellt wird.³

Der Wert dieser zum Teil doch recht aufwändig generierten Daten, man denke an Interviews oder Gesprächsnotizen, war zwar einer kleinen Gruppe an Forschenden bewusst, aber erst die Globalisierung und Digitalisierung, also die technischen Möglichkeiten, die das Internet bietet, macht es möglich, dass diese Materialien weltweit genutzt werden können. Um diese Möglichkeiten aber auch wirklich ausschöpfen zu können, müssen bestimmte Vorgaben erfüllt sein. Diese Anforderungen werden in den FAIR-Prinzipien erklärt. In vielen Fällen sind diese jedoch bis zum Einreichen eines eigenen Forschungsprojektes unbekannt und es existiert selten ein Plan, wie mit den Forschungsergebnissen nachhaltig und rechtssicher umgegangen werden soll. Wenn wir uns die FAIR-Prinzipien genauer ansehen, erkennen wir bei F (findable), das die Auffindbarkeit repräsentiert, die Forderung nach persistenten Identifiern für die Daten aber auch nach reichhaltigen Metadaten und die Wichtigkeit eines nachhaltigen Repositoriums.. Auch hier ergeben sich wieder neue Fragen. So ist der Begriff „Repositorium“ für viele Forschende neu, bzw. wird unterschiedlich interpretiert. Auch persistente Identifier und die Frage, wie man sie für den Forschungsoutput erhält, bedarf einiger Erklärungen. Die Frage nach den Metadaten ist ebenfalls nicht trivial. Einerseits müssen die Forschenden im Datenmanagementplan beschreiben, welches Schema sie verwenden werden, was für viele schwierig zu entscheiden ist, andererseits sollen die Daten mit einer hohen Expertise ausführlich beschrieben werden, was nicht nur Fachwissen sondern auch Zeit und damit Geld kostet.

Die Zugänglichkeit (Accessibility) mit A gekennzeichnet, verweist auf die Offenheit und Standardisierung von Protokollen und vor allem auch auf die Forderung, dass die Metadaten auch dann erreichbar sein müssen, wenn die Daten aus irgendeinem Grund gesperrt sein sollten, bzw. gesperrt werden müssen. Die Beschreibung der Daten ist also ein wesentlicher Faktor in der Forderung nach Transparenz. Die Offenlegung des Forschungsprozesses ist für Geisteswissenschaftler*innen nicht ganz so vertraut wie für Forschende aus Fächern, die es gewohnt sind jeden Schritt ihrer Tätigkeiten penibel festzuhalten und diese Laborbücher auch anderen zugänglich zu machen. Im Allgemeinen finden beispielsweise Text- und Bildinterpretationen eher alleine oder in kleineren Gruppen statt, selten werden die Ergebnisse vor einer Publikation öffentlich diskutiert. Vergleichen lässt sich dies beispielsweise mit der Entwicklung des Peer-Reviews, das erst nach und nach von den Geisteswissenschaften übernommen wurde.

3 Siehe etwa die Bibliothek samt beilagen des Germanisten Wendelin Schmidt-Dengler: <https://bibliothek.univie.ac.at/fb-germanistik/bibliothek_schmidt-dengler.html> (5.7.2021)

Die Interoperabilität, also die Austauschbarkeit von Daten und Metadaten, durch das I in FAIR repräsentiert, verlangt standardisierte Vokabularien und Metadaten, die miteinander verlinkt sind. Die Standardisierung von Metadaten ist in vielen Fächern jedoch noch ziemlich am Anfang. Vor allem in Fächern, die sehr nationale Begriffe verwenden, die in politisch brisanten Gebieten forschen, können schwer international geltende Standards erarbeitet werden. In einigen Fächern gibt es derzeit deshalb Bestrebungen, internationale Thesauri aufzubauen.

Der letzte Buchstabe, das R, verweist auf Re-Usability, also die Wiederverwendung der Daten. Auch hier spielen die Metadaten eine große Rolle, je besser und mehrsprachiger die Objekte beschrieben sind, desto leichter können Objekte nachgenutzt werden. Natürlich spielt auch die Wahl einer möglichst offenen Lizenz eine große Rolle. Die Offenlegung der Herkunft, der Provenienz der Daten, trägt ebenfalls dazu bei, dass Forschungsergebnisse – auch von anderen Disziplinen – nachgenutzt werden können. Nicht immer lassen sich jedoch diese Angaben in Repositorien abbilden.⁴

Den Forderungen nach den FAIR-Prinzipien werden je nach Fach unterschiedlich nachgekommen. Das Pilotprojekt „Ethnographische Datenarchivierung“ der DLE Bibliotheks- und Archivwesen am Institut für Kultur- und Sozialanthropologie der Universität Wien⁵ setzt sich seit 2017 mit diesen und anderen Herausforderungen auseinander. In diesem Projekt sollten in Abstimmung mit analogen universitären, nationalen und internationalen Initiativen Grundlagen für eine Datenmanagement-Strategie geschaffen werden, die den Anforderungen ethnographischer Forschung angepasst ist. Diese Erfahrungen werden mit anderen Fächern geteilt, die ähnliche Probleme haben.⁶

Viele Forschende schrecken jedoch nach wie vor vor diesen Anforderungen zunächst einmal zurück und verwechseln oft die Forderung nach FAIR mit der Forderung nach völliger Offenheit. FAIR bedeutet jedoch nicht, dass alle Daten offen und frei verfügbar gemacht werden müssen, das wäre aus rechtlichen und ethischen Gründen gar nicht möglich. Aber die Daten sollen so transparent wie möglich beschrieben werden und diese Metadaten sollen frei verfügbar und zugänglich sein. Der Aspekt der Maschinenlesbarkeit tritt dabei immer stärker in den Vordergrund. Auch hier besteht für die Geisteswissenschaftler*innen kein Grund zur Sorge, denn ein geeignetes Repository, das allerdings sorgfältig ausgewählt werden sollte, ermöglicht, dass diesen Forderungen entsprochen wird. Wichtig ist es, den Forschenden die Gründe, die hinter diesen für sie zum Teil doch recht neuen Forderungen stehen, zu erklären und ihnen Lösungsmöglichkeiten anzubieten. Die Unsicherheit ist oft darin begründet, dass die Jungforscher*innen während des Studiums kaum mit Datenmanagement in Berührung kommen. Viele Betreuer*innen der Abschlussarbeiten wissen ebenfalls zu wenig darüber. Das führt nicht nur dazu, dass viele Informationen, die von anderen Forscher*innen weitergenutzt werden könnten, verloren gehen, sondern auch zu Rechtsunsicherheiten und im schlimmsten Fall zu juristischen Problemen. Vor allem bei bildlastigen Studienrichtungen wie beispielsweise der

4 PHAIDRA ermöglicht diese Funktion bereits seit Beginn. Hier gibt es zahlreiche Metadatenfelder, die auf die Herkunft der Daten verweisen und zusätzlich auch die Möglichkeit schaffen, Urheber*innen der Originale einzutragen.

5 <<http://eda.univie.ac.at>> (5.7.2021).

6 Eberhard, I. und Kraus, W. (2018) „Der Elefant im Raum. Ethnographisches Forschungsdatenmanagement als Herausforderung für Repositorien“, Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare, 71(1), S. 41–52. <<https://doi.org/10.31263/voebm.v71i1.2018>>.

Kunstgeschichte, bei der oft die Beschreibung einer Abbildung im Mittelpunkt steht, sind die Rechte für den Abdruck, bzw. die Archivierung in einem Repositorium vorab zu klären.

Ethische Aspekte

In Datenmanagementplänen wird oft auch dezidiert nach ethischen Aspekten der Forschung gefragt. Beim FWF heißt es beispielsweise: „Gibt es ethische Hindernisse, um alle Forschungsdaten vollständig oder nur teilweise zugänglich zu machen? Falls zutreffend: Wie planen Sie den Umgang mit sensiblen Daten während und nach dem Projekt?“⁷ Aufgrund ihrer Fachexpertise sind hier die meisten Forschenden gut informiert darüber, wie sie mit Daten, die aus irgendeinem Grund heikel sein könnten, umgehen. Wenn es allerdings um Fragen der Archivierung, bzw. Langzeitarchivierung geht, benötigen viele von Ihnen nicht nur gezielte Hinweise, sondern auch konkrete Tools um diese Art der Daten adäquat zu managen. Anonymisierungstools, geeignete und sichere Speicherplätze werden hier gerne angenommen. Weniger diskutiert werden allerdings Fragen, wie man mit politisch brisanten Daten umgehen soll. Es fehlen beispielsweise Lösungen, wie man mit Abbildungen oder Texten aus der Zeit des Nationalsozialismus umgehen soll, die man zwar archivieren und für wissenschaftliche Zwecke abrufbar machen möchte, die aber gleichzeitig kontextualisiert und mit entsprechenden Beschreibungen und Erklärungen versehen werden müssen. Hier den richtigen Weg zwischen aus Furcht vor juristischen Konsequenzen zurückgehaltenen Informationen und Open Science zu finden ist eine Gratwanderung und bedarf genauer Anleitungen und viel Hintergrundwissen. Zu diesem Thema entstehen an der Universität Wien gerade mehrere Arbeitsgruppen, die sich dem Thema nähern. Sich dieser Thematik anzunähern ist auch ein ganz guter konstruktiver Ansatz, wie man als Forschungsdatenmanager*in an Bibliotheken am besten Geisteswissenschaftler*innen dabei hilft „as open as possible, as closed as necessary“ zu agieren.

Bei all den Anfragen, die wir im Bereich Datenmanagement erhalten, sind die Geisteswissenschaften recht weit vorne dabei. Unser Anliegen ist es natürlich alle Forschenden möglichst gut zu unterstützen und zu beraten. Die Forschungsunterstützenden Services der UB Wien⁸ geben hier größtmöglichen Input entlang des gesamten Forschungsprozesses. Der Zentrale Informatikdienst stellt zahlreiche Tools für das Datenmanagement zur Verfügung, unter anderem eine Cloudlösung, mit der Daten geteilt werden können und auch GitLAB und Shares.

Strategien

Je intensiver wir mit den Forschenden zusammenarbeiten, je mehr wir von den einzelnen Projekten erfahren, desto differenzierter können auch wir auf Fragen eingehen. Wichtig erscheinen folgende Voraussetzungen:

7 <https://www.fwf.ac.at/fileadmin/files/Dokumente/Open_Access/FWF_DMPTemplate_d.pdf>.

8 <<https://bibliothek.univie.ac.at/forschungsunterstuetzung/>> (5.7.2021).

- **Nähe zu den Forschenden**

Um herauszufinden, wer welche Unterstützung in welchem Umfang benötigt, führen wir Umfragen, Interviews und Podcasts durch und sprechen bei Schulungen, Besprechungen und Vorträgen mit den Wissenschaftler*innen über ihren Umgang mit Forschungsdaten. Außerdem findet – im Kooperation mit dem Projekt „FAIR Data Austria“⁹ eine sehr gut angenommene Webinarreihe¹⁰ unter dem Titel „Forschungsdatenmanagement in Österreich“ statt. Durch diese Kontakte erfahren wir viel über laufende und geplante Forschungen, wir lernen Methoden und Tools kennen, vor allem auch durch die erforderlichen Datenmanagementpläne, und wir erfahren auch viel über die Quellen mit denen die Forschenden arbeiten.

- **Vernetzung**

Repositorienmanager*innen und Forschungsdatenmanager*innen vernetzen sich um sich über Erfahrungen auszutauschen, wie etwas im Netzwerk für Repositorienmanager*innen (RepManNet)¹¹ in diversen Projekten, etwa FAIR Data Austria oder auch international, wie beispielsweise auch in COAR¹², der Confederation of Open Access Repositories, bzw. der RDA Research Data Alliance¹³, im Rahmen der European Open Science Cloud (EOSC)¹⁴ und den diversen Arbeitsgruppen. So lernen wir viel von anderen Strategien und geben zugleich auch unser Wissen weiter.

Wir vernetzen aber auch die Forschenden untereinander und geben Ihnen damit die Möglichkeit fächerübergreifend über ähnliche Herausforderungen miteinander zu diskutieren und eventuell gemeinsame Strategien an der Universität Wien zu entwickeln. Das Netzwerk für nachhaltige Softwareentwicklung „Code4research“ unterstützt Forschende beim Aufbau von technischen Infrastrukturen.¹⁵

Wie wichtig die Vernetzung der Forschungsdatenmanager*innen mit den Forschenden ist um ganz genau zu verstehen welche Anforderungen und Bedürfnisse es gibt, zeigt das Projekt „Ethnographische Datenarchivierung“ an der Universität Wien, wie Eberhard und Kraus schreiben: „Durch die intensive Zusammenarbeit mit dem PHAIDRA-Team kann das Projekt EDA jedoch auf die Weiterentwicklung dieses Digital Asset Management Systems direkten Einfluss nehmen. Somit ergibt sich aus dieser Zusammenarbeit – eine entsprechende personelle und finanzielle Ausstattung vorausgesetzt – die Chance, hier gemeinsam Standards zu setzen.“¹⁶

9 <<https://forschungsdaten.at/fda/>> (5.7.2021).

10 <<https://datamanagement.univie.ac.at/forschungsdatenmanagement/veranstaltungsreihe-forschungsdatenmanagement/>> (5.7.2021).

11 <<https://datamanagement.univie.ac.at/forschungsdatenmanagement/netzwerk-fuer-repositorienmanagerinnen-repmanet/>> (5.7.2021) Siehe auch: Blumesberger, Susanne: Das RepManNet – das österreichische Netzwerk für RepositorienmanagerInnen. In: Mitteilungen der VÖB. Bd. 72, Nr. 1 (2019). S. 160–163 <<https://doi.org/10.31263/voebm.v72i1.2291>>.

12 <<https://www.coar-repositories.org/>> (5.7.2021).

13 <<https://rd-alliance.org/>> (5.7.2021).

14 <<https://eosc-portal.eu/>> (5.7.2021).

15 <<https://datamanagement.univie.ac.at/forschungsdatenmanagement/netzwerk-nachhaltige-softwareentwicklung/>> (5.7.2021).

16 Eberhad/Kraus 2018, S. 50.

- **Konkrete Tools**

Wir bieten unterschiedliche Tools für das Datenmanagement an, während und nach dem Forschungsprojekt können Daten unterschiedlicher Art bearbeitet und archiviert werden. Wir sind beispielsweise dem Konsortium Transkribus¹⁷ beigetreten, eine Software für die automatische Erkennung und Transkription von Handschriften. Aber auch elektronische Notebooks und verschiedene Speichermöglichkeiten werden angeboten.

- **Verbesserte Services durch Data Stewards**

Mit dem Aufbau von Data Stewards sollen in Zukunft Wissenschaftler*innen entlastet werden. Data Stewards helfen disziplinspezifisch den Forschenden beim Managen ihrer Daten. Aus den jeweiligen Disziplinen stammend wissen sie genau welche Anforderungen hier bestehen, kennen aber auch die forschungsunterstützenden Services der Bibliothek sehr gut, sind technisch am Laufenden und können die Forschenden somit zeitnah und punktgenau unterstützen.

Fazit

Was benötigen forschungsunterstützende Services in Zukunft?

Aufbau von Wissen über Daten und Datenmanagement bereits während des Studiums, vermehrt auch technisches und juristisches Wissen werden benötigt. Ebenso wären Diskussionen über ethische Aspekte in der Forschung und vor allem beim Umgang mit Forschungsdaten wichtig. Dazu benötigen wir auch geeignete Ausbildungsmöglichkeiten, die uns begleiten, die regelmäßig evaluiert werden und die möglichst flexibel genutzt werden können. Technisches Basiswissen wird immer wichtiger, denn nur Mitarbeiter*innen, die über technische Grundkenntnisse verfügen, können eine umfassende Beratung im Bereich Datenmanagement anbieten. Nicht zuletzt benötigen wir eine besser ausgebaute personelle und technische Infrastruktur um für den Zuwachs an Anfragen im Bereich Datenmanagement und Datenmanagementplanberatung gut vorbereitet zu sein.

17 <<https://transkribus.eu/lite/de>> (5.7.2021).

Literatur

- Blumesberger, Susanne: Forschungsdatenmanagement gestern, heute und morgen zwischen FAIR, CARE und EOSC. Ein Praxisbericht der Universität Wien. In: b.i.t. online, 23 (2020) Nr. 5. S. 500–508.
- Blumesberger, Susanne: The data management plan - the common thread through the project. DMP consulting at the University of Vienna. In: Proceedings of Science, Austrian Citizen Science Conference 2020 – ACSC2020, 2021.
- Blumesberger, Susanne: Das RepManNet – das österreichische Netzwerk für RepositorienmanagerInnen. In : Mitteilungen der VÖB. Bd. 72, Nr. 1 (2019). S. 160–163 <<https://doi.org/10.31263/voebm.v72i1.2291>>.
- Eberhard, I. und Kraus, W. (2018) „Der Elefant im Raum. Ethnographisches Forschungsdatenmanagement als Herausforderung für Repositorien“, Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare, 71(1), S. 41–52. <<https://doi.org/10.31263/voebm.v71i1.2018>>.