



Datenmanagementplan (DMP) für die Abschlussarbeit¹ Beispiel

Titel der Arbeit: Nutzer*innen Studie zur Sprachvielfalt in den Beständen der Berliner Bibliotheken
Art der Arbeit: Masterarbeit
Name: Kim Musterperson
Matrikelnummer: 111111
Kontakt: kim.musterperson@hu-berlin.de /
kim.musterperson@gmx.de
Datum Einreichung: 29.02.2022

Inhalt

1. Allgemein	2
2. Inhaltliche Einordnung.....	2
3. Technische Einordnung	3
4. Datennutzung	4
5. Metadaten und Referenzierung	5
6. Rechtliche und ethische Fragen	6
7. Speicherung und Langzeitarchivierung.....	7

Generelle Anmerkungen:

- Wenn Sie keine Anmerkungen zu einer Frage haben oder diese nicht auf Ihre Abschlussarbeit zutrifft, vermerken Sie es bitte (z. B. „trifft nicht zu“).
- Wenn Sie mehrere Datensatztypen (z. B. Interviewaufnahmen und deren Transkriptionen) oder mehrere Datensammlungen aus z. B. unterschiedlichen Quellen haben, beschreiben Sie diese bitte einzeln in dem jeweiligen Kapitel, wo sinnvoll.
- Sie können dem DMP gerne Anhänge zufügen (bitte nummerieren und darauf dann verweisen) oder auf entsprechende Kapitel in Ihrer Abschlussarbeit verweisen.
- Bitte beachten Sie auch die jeweilige FDM-Policy, Richtlinien und/oder Handreichungen der Humboldt-Universität (HU).

¹ Dieses DMP Template wurde für den gemeinsamen Masterstudiengang Digitales Datenmanagement (DDM) des Instituts für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin und des Fachbereichs Informationswissenschaften der Fachhochschule Potsdam entwickelt. <https://www.ddm-master.de/>



1. Allgemein

1.1 Thema

1.1.1. Wie lautet die primäre Forschungsfrage der Abschlussarbeit?

Die Arbeit möchte untersuchen, ob die Bestände der Berliner Bibliotheken der Sprachvielfalt der Nutzer*innen gerecht wird. Zu diesem Zweck werden Interviews mit aktiven Bibliotheksbesucher*innen durchgeführt. Die entstehenden Interviewaufzeichnungen werden transkribiert, analysiert und ausgewertet. Die Datensammlung soll dazu dienen, einen Überblick über die sprachlichen Bedürfnisse der Bibliotheksnutzenden zu erhalten und untersuchen, ob diese sich mit dem aktuellen Angebot der Berliner Bibliotheken decken.

Die Forschungsfrage lautet: *Werden die Bestände der Berliner Bibliotheken der Sprachvielfalt ihrer Nutzenden gerecht?*

1.1.2. Bitte geben Sie einige Schlagwörter² zur Forschungsfrage bzw. Fragestellung an.

GND-Schlagwörter: Bibliothek (nid=4006439-6), Bibliothekswesen (idn=1021967998), Sprachwandel (nid=4056508-7), Stadtbevölkerung (nid=4138964-5)

1.1.3. Welchen Regeln oder Richtlinien (HU) zum Umgang mit den in der Abschlussarbeit erhobenen Forschungsdaten folgen Sie für den DMP? Bitte referenzieren Sie diese hier inklusive Version bzw. Veröffentlichungsjahr.

Grundsätze zum Umgang mit Forschungsdaten an der Humboldt-Universität zu Berlin (Forschungsdaten-Policy) vom 8. Juli 2014 <https://www.cms.hu-berlin.de/de/dl/dataman/infos/policy>

2. Inhaltliche Einordnung

NB: Bitte beschreiben Sie jeden Datensatztyp oder Datensammlung einzeln in dem jeweiligen Kapitel, wo sinnvoll.

2.1. Datensatz

2.1.1 Um welche Arten von Daten handelt es sich? Bitte in wenigen Zeilen kurz beschreiben.

Für die Arbeit wurden 10 Interviews mit der Videokonferenz Software Zoom durchgeführt. Es liegen Audio-, so wie Video-Dateien der Interviewaufnahmen vor.

² Hier gern auch ein anerkanntes (Fach)Vokabular nutzen, die Dewey Dezimalklassifikation (DDC deutsch, <https://deweysearchde.pansoft.de/webdeweysearch/mainClasses.html?catalogs=DNB>) oder Gemeinsame Normdatei (GND, https://gnd.network/Webs/gnd/DE/Home/home_node.html) bieten sich für fachübergreifende Terme an. Bitte angeben, ob und falls ja, welches Vokabular benutzt wurde bzw. ob freie Schlagwortvergabe angewendet wird.

Anschließend wurden die Transkripte mit Hilfe der Software für qualitative Datenanalyse MAXQDA erstellt, codiert und analysiert.

2.2 Datenursprung

2.2.1 Werden die Daten selbst erzeugt oder nachgenutzt?

Die Daten wurden selbst erhoben.

2.2.2 Wenn die Daten nachgenutzt werden, wer hat die Daten erzeugt? Bitte mit Angabe des Identifiers, falls vorhanden, z.B. DOI³.

Trifft nicht zu, siehe oben.

2.3. Reproduzierbarkeit

2.3.1 Sind die Daten reproduzierbar, d. h. ließen sie sich, wenn sie verloren gingen, erneut erstellen oder erheben?

Die Daten sind nicht exakt reproduzierbar. Da es sich um Interviews handelt, würde es bestimmte Abweichungen zwischen den ersten und zweiten Interviews geben.

2.4 Nachnutzung

2.4.1 Für welche Personen, Gruppen oder Institutionen könnte dieser Datensatz (für die Nachnutzung) von Interesse sein? Für welche Szenarien ist dies denkbar?

Der Datensatz könnte abgesehen von der Bibliotheks- und Informationswissenschaft für die Sozialwissenschaft von Interesse sein, da die Arbeit die Sprachvielfalt und den Sprachwandel in einer Stadtgesellschaft untersucht. Des Weiteren könnte die Untersuchung für Menschen, die sich für die Geschichte und gesellschaftliche Entwicklung Berlins interessieren. Außerdem könnte die Arbeit allgemein für das Bibliothekswesen und eventuell für die Politik, von Interesse sein.

3. Technische Einordnung

3.1 Datenerhebung

3.1.1 Wann erfolgt(e) die Erhebung bzw. Erstellung der Daten?

Die Interviews wurden vom 28.07.21 – 09.08.21 durchgeführt und bis Ende August transkribiert.

³ DOI, <https://www.doi.org/>

3.1.2 Wann erfolgt(e) die Datenbereinigung / -aufbereitung bzw. Datenanalyse?

Die Datenanalyse fand im September 2021 statt.

3.1 Datengröße

3.1.1 Was ist die tatsächliche oder erwartete Größe der Daten(typen)?

Die Audio-/Videodateien liegen im Format MP4 vor. Jedes Interview benötigt ca. 100 Megabyte. Somit liegen wir bei 10 Interviews bei einer Größe von ca. 1 Gigabyte. Die Transkripte der Interviews liegen als PDF-Dateien vor, mit 5 Megabyte. Die MAXQDA-Datei hat eine Größe von 3 Megabyte.

3.2 Formate

3.2.1 In welchen Formaten⁴ liegen die Daten vor?

Die Interviews liegen in folgenden Datei-Formaten vor: Audio-/Video-Dateien (MP4), die Transkripte als PDF-Dateien und die Analyse als MAXQDA-Datei.

3.3 Werkzeuge

3.3.1 Welche Instrumente, Software, Technologien oder Verfahren werden zur Erzeugung, Erfassung, Bereinigung, Analyse und/oder Visualisierung der Daten genutzt? Bitte (falls möglich) mit Versionsnummer und Referenz in Form einer Adresse jeweils angeben.

Die Aufnahmen wurden mit der Videokonferenz Software Zoom (Version 5.4.3) aufgenommen. Um die Transkripte zu erstellen, so wie die Codierung und Analyse dieser, wurde die Software MAXQDA2020 Version verwendet. Außerdem liegen die Interview-Abschriften auch als PDF-Dateien vor.

3.3.2 Welche Software, Verfahren oder Technologien sind notwendig, um die Daten zu nutzen?

Eine Audio-/Video-Wiedergabe-Software, ein PDF-Viewer und MAXQDA.

3.4 Versionierung

3.4.1 Werden verschiedene Versionen der Daten erzeugt (z. B. durch verschiedene Weiterbearbeitungsprozesse bzw. Bereinigung von Daten)?

Es liegen die Rohdaten in Form von Audio-/Video-Aufnahmen der Interviews vor, genauso wie die Abschriften davon. In MAXQDA werden die Daten dann codiert und analysiert.

⁴ Vgl. z. B. DROID zur Format-Erkennung, <http://digital-preservation.github.io/droid/>

4. Datennutzung

4.1 Datenorganisation

4.1.1 Gibt es eine Strategie zur Benennung der Daten? Wenn ja, bitte skizzieren Sie sie kurz.

Die Daten werden mit dem Namen „Person“ und einer Zahl versehen, die der chronologischen Aufzeichnung der Interviews folgt. Außerdem wird das Datum, an dem das Interview durchgeführt wurde, angegeben. Beispiel: Interview_Person1_29_07_21

4.2 Datenspeicherung und -sicherheit

4.2.1 Wer darf (zukünftig) auf die Daten zugreifen?

Die Gutachter*innen dürfen auf die Daten zugreifen. Die anonymisierten Daten können auch veröffentlicht und von anderen verwendet werden.

4.2.2 Wie und wie oft werden Backups der Daten erstellt?

Es wird jede Woche ein Backup erstellt und die Daten werden auf einer externen Festplatte gespeichert. Außerdem liegen die Daten zusätzlich in der HU-Box und werden dort täglich aktualisiert.

4.3 Interoperabilität

4.3.1 Sind die Datenformate im Sinne der FAIR-Prinzipien interoperabel, d.h. geeignet für den Datenaustausch und die Nachnutzung zwischen bzw. von unterschiedlichen Forschenden, Institutionen, Organisationen und Ländern?

Für die vorliegende Untersuchung wurden teilweise proprietäre Datenformate verwendet, wie MAXQDA. Die Interview-Transkripte liegen allerdings im offenen Format PDF vor. Die Audio-/Videoaufnahmen liegen im MP4 Format vor, welches ein internationaler Standard ist.

4.4 Weitergabe und Veröffentlichung

4.4.1 Ist es geplant, die Daten nach Abgabe der Abschlussarbeit zu veröffentlichen oder zu teilen?

Ja, die Forschungsdaten sollen auf dem edoc-Server: Open-Access-Publikationsserver der Humboldt-Universität veröffentlicht werden. <https://edoc.hu-berlin.de/>

4.4.2 Wenn nicht, skizzieren Sie kurz rechtliche und/oder vertragliche Gründe und freiwillige Einschränkungen.

Siehe oben.

4.4.3 Wenn ja, unter welchen Nutzungsbedingungen oder welcher Lizenz sollen die Daten veröffentlicht bzw. geteilt werden?

Die Daten sollen unter der Creative Commons Lizenz CC-BY 3.0 DE (<https://creativecommons.org/licenses/?lang=de>) veröffentlicht werden.

4.5 Qualitätssicherung

4.5.1 Welche Maßnahmen zur Qualitätssicherung (z. B. Plausibilitätsprüfung von Datenwerten) werden für die Daten ergriffen?

Da es sich bei dieser Arbeit um Daten aus Interviews handelt, ist es schwierig diese Datenwerte einer Plausibilitätsprüfung zu unterziehen. Die Interviews und deren Bearbeitung wurden nach bestem Wissen und Gewissen durchgeführt. Es wurde darauf geachtet die Regeln der guten wissenschaftlichen Praxis⁵ einzuhalten. Wie die Interviews genau vorbereitet, durchgeführt und anschließend bearbeitet wurden wird ausführlich im Methodenkapitel der Masterarbeit beschrieben.

4.6 Datenintegration

4.6.1 Falls Daten aus verschiedenen Quellen (z. B. Anpassung Skalierung, Zeiträume, Ortsangaben) integriert werden, wie wird dies gewährleistet?

Trifft nicht zu. Die Daten liegen in unterschiedlichen Formaten vor und werden auch so eingereicht.

5. Metadaten und Referenzierung

5.1 Metadaten

5.1.1 Welche Informationen sind für Außenstehende notwendig, um die Daten zu verstehen (d. h. ihre Erhebung bzw. Entstehung, Analyse sowie die auf ihrer Basis gewonnenen Forschungsergebnisse nachvollziehen) und nachnutzen zu können?

Der Interview-Leitfaden, der den vorliegenden Daten zu Grunde liegt befindet sich im Anhang. Außerdem sind die Informationen im Methodenkapitel der Masterarbeit zu finden.

5.1.2 Welche Standards, Ontologien, Klassifikationen etc. werden zur Beschreibung der Daten und Kontextinformation genutzt?

Da es sich bei den vorliegenden Daten um Interviews handelt, werden keine Standards, Ontologien oder Klassifikationen genutzt. Es wurde ein ausführliches Codebuch für die codierten Interviews angelegt, dieses wurde induktiv aus den Daten heraus erstellt. Das

⁵ www.amb.hu-berlin.de/2014/06/06_2014_20140130%20Beschlussversion%20Satzung%20Wissenschaftliches%20Fehlverhalten_DRUCK.pdf

Codebuch (kontrollierte Vokabularliste) befindet sich ebenfalls hier im Anhang und im Anhang der schriftlichen Arbeit als PDF-Datei.

6. Rechtliche und ethische Fragen

6.1 Personenbezogene Daten

6.1.1 Enthalten die Daten personenbezogene Informationen?

Nein, es werden keine personenbezogenen Informationen abgefragt. Informationen, wie zum Beispiel das Alter der interviewten Person oder ihre Muttersprache, kann nicht auf die Person zurückgeführt werden. Die Namen der interviewten Personen werden anonymisiert.

6.2 Sensible Daten

6.2.1 Enthalten die Forschungsdaten besondere Kategorien personenbezogener Daten nach Artikel 9 der DSGVO ("Angaben über die rassische und ethnische Herkunft, politische Meinungen, religiöse oder philosophische Überzeugungen, Gewerkschaftszugehörigkeit, Gesundheit oder Sexualleben")⁶?

Es sind Daten enthalten zur Herkunft der interviewten Personen und der Sprachen, die sie sprechen.

6.2.2 Werden die Daten anonymisiert oder pseudonymisiert?

Die Namen der interviewten Personen werden anonymisiert. In der schriftlichen Arbeit werden Begriffe wie Person, Interviewte*r oder Teilnehmer*in verwendet.

6.2.3 Haben Sie eine "informierte Einwilligung" der Betroffenen eingeholt? Fügen Sie bitte ein Template der Einverständniserklärung als Anlage bei.

Die Einverständniserklärung befindet sich im Anhang. Die Einverständniserklärung besteht aus einem Informationsschreiben, Hinweis zum Datenschutz und der Einwilligungserklärung. Es wird festgelegt, dass die erhobenen Interview-Daten nur anonymisiert verwendet werden. Außerdem willigten die Personen ein, dass die anonymisierte Abschrift des Interviews als Teil der Forschungsdaten auf dem edoc-Server veröffentlicht, werden dürfen.

6.2.4 Wenn keine "informierte Einwilligung" eingeholt wird, begründen Sie dies bitte.

Trifft nicht zu, siehe oben.

6.2.5 Wo und wie sind die "informierten Einwilligungen" abgelegt?

⁶ <https://dsgvo-gesetz.de/art-9-dsgvo/>

Die Einverständniserklärungen befinden sich mit allen anderen erhobenen Daten auf einer externen Festplatte in den privaten Räumen der Autorin und außerdem in der HU-Box.

6.2.6 Bis wann werden die (un-anonymisierten bzw. un-pseudonymisierten) Originaldaten spätestens sicher vernichtet?

Die Audio-/Video-Daten werden nach erfolgreichem Bestehen der Masterarbeit gelöscht.

6.3 Urheber- oder verwandte Schutzrechte

6.3.1 Werden Daten genutzt und/oder erstellt, die durch Urheber- oder verwandte Schutzrechte geschützt sind?

Trifft nicht zu.

7. Speicherung und Langzeitarchivierung

7.1 Wo werden die Daten (einschließlich Metadaten, Dokumentation und ggf. relevantem Code bzw. relevanter Software) während Phase der Erarbeitung der Abschlussarbeit gespeichert?

Die Daten werden zum einen in der HU-Box der Autorin gespeichert und zum anderen auf einer externen Festplatte, die sich in der Wohnung der Autorin befindet.

7.2 Wo werden die Daten (einschließlich Metadaten, Dokumentation und ggf. relevantem Code bzw. relevanter Software) nach dem Ende der Abschlussarbeit gespeichert bzw. archiviert?

Auf der externen Festplatte der Autorin. Die anonymisierten Transkripte, sowie das Codebuch und die MAXQDA-Datei werden auf dem edoc-Server veröffentlicht.

7.3 Handelt es sich dabei um ein zertifiziertes Repositorium oder Datenzentrum (z.B. durch das CoreTrustSeal⁸, nestor-Siegel⁹ oder ISO 16363¹⁰)? (Wurden mehrere Langzeitarchivierungsoptionen ausgewählt, kann die Frage bejaht werden, wenn dies auf mindestens eine der Optionen zutrifft).

Ja, der edoc-Server ist DINI zertifiziert. Genaueres im Kapitel Nachhaltigkeit in den „Leitlinien für den edoc-Server (Open-Access-Publikationsserver der Humboldt-Universität zu Berlin“: https://edoc-info.hu-berlin.de/de/nutzung/nutzung_leitlinien

⁷ Vgl. z. B. Registry of Research Data Repositories, <https://www.re3data.org/>

⁸ <https://www.coretrustseal.org/about/>

⁹ https://www.langzeitarchivierung.de/Webs/nestor/DE/Zertifizierung/nestor_Siegel/nestor_siegel_node.html;jsessionid=766FAA5089CB1958F7EAE00D2BA51F3.internet542

¹⁰ http://ikeep.com/ISO_16363_ISO_16919_DIN_31644