

Dokumentation von Daten & Code in Abschlussarbeiten

Selbst erstellter Programmcode und erhobene Daten sind die wichtigen Teile Ihrer Abschlussarbeit. Sie müssen diese den Prüfer*innen und ggf. bei späterer Veröffentlichung in der Bibliothek auch den Leser*innen in nachvollziehbarer Form zur Verfügung stellen. Sie entscheiden jederzeit, wer die Daten/den Code neben den Prüfer*innen sehen darf. Besprechen Sie die Anforderungen in Ihrem konkreten Einzelfall immer mit Ihren Betreuer*innen!

1 Was muss dokumentiert werden?

Gute wissenschaftliche Praxis bedeutet, dass Ihre Forschungsergebnisse für Dritte nachvollziehbar und replizierbar sind. Deshalb werden Programmcode und (digitale) Daten, die während einer wissenschaftlichen Tätigkeit (z.B. durch Messungen, Befragungen, Experimente, Sammlung von Webinhalten) entstehen, dokumentiert.

Beispiele für Programmcode:

- Python, R oder SPSS-Code zur Datenauswertung
- Code zur Datensammlung
- Bei selbst erstellten KI Modellen: Code und Modelldatei

Beispiele für Forschungsdaten:

- Interviewleitfäden und Transkripte, Codebücher (z.B. bei qualitativen Inhaltsanalysen), Korrespondenz (z.B. E-Mail-Verkehr mit Expert*innen)
- Audio, Bild- oder Videodateien (z.B. Interview-Aufnahmen oder Fotos),
- Tabellen oder csv-Dateien, json-Dateien
- Internet-Daten (z.B. Tweets, Inhalte von Beiträgen, Likes, Kommentare etc.)
- (ausgefüllte) Fragebögen

2 Erstellung eines Datenmanagementplans (DMP)

Der Plan wird während der **Vorbereitung der Abschlussarbeit** angelegt und während der einzelnen Arbeitsphasen Ihrer Abschlussarbeit ergänzt.

[Checkliste Forschungsdaten](#)

Der Datenmanagementplan enthält folgende Abschnitte:

- Angaben zu Autor*innen, Betreuer*innen, Titel der Abschlussarbeit, Was und wie soll untersucht werden (kurz)
- geplante Datenerhebungsmethode, Anzuwendende Regeln oder Richtlinien (z.B. zu Datenschutz, Anonymisierungsmaßnahmen)
- Art (Dateiformat) und Umfang (Größe) der zu erhebenden Daten
- Geplanter Ablageort und geplante Ablagestruktur während des Projekts und nach dem Projektende (wo gespeichert, Dateibenennungskonventionen, Versionierung, Backup-Strategie, Langzeitarchivierung)

3 Veröffentlichung von Daten und Code

Die Veröffentlichung von Daten und Code muss spätestens vor **Abgabe der Abschlussarbeit** organisiert sein.

Wo: Sie können Ihre Daten öffentlich zugänglich machen oder nur für von ihnen ausgewählte Personen, z.B. den Prüfer*innen. Wählen Sie für die Veröffentlichung ein geeignetes Repository (z.B. GitLab oder Github; Hinweis: die im HTW-GitLab erstellten Repositories können nicht öffentlich zugänglich gemacht werden und werden bei Ende des Studiums gelöscht). Insbesondere wenn Ihr Datensatz groß ist (>500 Mbyte) besprechen Sie mit den Prüfer*innen, wie die Daten zur Verfügung gestellt werden sollen.

Lizenz: Wählen Sie für Sie geeignete Lizenzen und geben Sie diese explizit an. In den meisten Fällen werden die Daten und die Software unterschiedliche Lizenzen benötigen. Tipp: Hilfreiches Tool für die Entscheidung zu einer Lizenz: <https://ufal.github.io/public-license-selector/>

Dokumentation: Dokumentieren Sie Daten und Code in einer README.

Inhalt der Readme:

<u>Datensatz/Softwarepaket (Was?)</u> <ul style="list-style-type: none">• Arten von Daten• Sprache• Genutzte wissenschaftliche Methode	W-Fragen: Was für Daten wurden wofür und wie erhoben? Bzw. Was für Software wurde wofür erstellt? Beispiel Daten: Es wurden 19 semi-strukturierte Expert*innen Interviews zum Thema xx mit der Videokonferenz-Software yy durchgeführt und aufgezeichnet. Die Dateien wurden transkribiert und im Rahmen des Ansatzes der <i>Grounded Theory</i> analysiert. Sprache: Deutsch. Beispiel Software: Es wurden Python Skripte für eine statistische Datenanalyse und für die Erzeugung von Surrogatdaten erstellt. Sprache: DE, EN
<u>Datenursprung (Wer?)</u> <ul style="list-style-type: none">• Author*innen,• Lizenz• Identifier, z.B. DOI	Beispiel 1: Die Daten wurden selbst erhoben. Beispiel 2: Offene Datensammlung: van der Burg, S. (2024). Dollar street 10 - 64x64x3 [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10970014 Beispiel 3: Daten mit eingeschränktem Zugriff: beschreiben Sie wer die Daten unter welchen Voraussetzungen erhalten kann.
<u>Zeitraum (Wann?)</u>	Zeitraum der Datenerhebung/-bereinigung/-aufbereitung/-analyse, Softwareentwicklung, Beispiel: Die Interviews wurden am 25.07.2024 bis 07.08.2024 durchgeführt und bis Mitte August transkribiert. Die Datenanalyse fand im September 2024 statt.

<u>Datenformate und -größe</u> (Welche?, Wie viel?)	Beispiel: Interviews liegen in folgenden Formaten vor: <ul style="list-style-type: none"> • Video-Dateien (.mp4): 100 MB pro Interview, • die Transkripte (.pdf): 5 MB (alle zusammen), • Analyse als QualCoder Datei: 3 MB
<u>Werkzeuge</u> (zur Erzeugung, Erfassung, Bereinigung, Analyse, Visualisierung), inkl. Versionsnummer, Referenz (Link)	Listen Sie bitte hier nur das, was sie tatsächlich genutzt haben. Beispiel: <ul style="list-style-type: none"> • Die Aufnahmen: Videokonferenz Software Zoom (Ver. 5.4.3) https://www.zoom.com/de • Transkription: f4x Software https://www.audiotranskription.de/ • Analyse: QualCoder Ver. 3.6 https://github.com/ccbogel/QualCoder/wiki
<u>Qualitätssicherung</u>	Schritte zur Sicherung der Qualität und Integrität der aufbereiteten Daten und Analyseergebnisse Beispiel: Manuelle Prüfung und Korrektur der Transkripte. Entfernung von Außereißern und Ersetzen fehlender Werte während der Datenbereinigung. Python-Code für die Datenbereinigung unter [Respository].
<u>Datenschutz</u> <ul style="list-style-type: none"> • sensible/ personenbezogene Datenanalyse • Urheber- oder andere Schutzrechte 	Bei den sensiblen/personenbezogenen Daten: <ul style="list-style-type: none"> • Einwilligungserklärungen der Betroffenen • Pseudo-/Anonymisierungsmaßnahmen Infos zur Urheberrecht: https://bibliothek.htw-berlin.de/urheberrecht/ Beispiel: Ergebnisse aus der Interview Analyse wurden mit Amnesia (https://amnesia.openaire.eu/) anonymisiert.
<u>Ablageort</u> <ul style="list-style-type: none"> • Ordnerstruktur • Dateibenennungskonventionen • Ggf. zusätzliche Informationen zum Zugriff (Passwörter, zugriffsberechtigte Personen) • alternative Ablageorte • Löschdatum 	Beispiel 1: Die Rohdaten sind unter folgendem DOI abrufbar: https://doi.org/10.5281/zenodo.10970014 Beispiel 2: Die Daten sollen nach 10 Jahren Aufbewahrungszeit gelöscht werden: Juni 2036.

Weitere Informationen finden Sie unter <https://forschungsdaten.info/> sowie unter <https://www.ibi.hu-berlin.de/de/studium/rundumdasstudium/fdm-fuer-studierende>