

Konzept zum Umgang mit Forschungsdaten¹

Datenbeschreibung: Die im Projekt XYZ zu erhebenden Forschungsdaten werden über einen Onlinefragenbogen gesammelt. Hierzu wird die vom Computer- und Medienservice (CMS) der Humboldt-Universität zu Berlin bereitgestellte Software LimeSurvey genutzt. Die Analyse der Umfragedaten erfolgt mit Hilfe der Open Source Statistik-Software R und wird in Form des Datensatzes (csv), des R-Analyseskriptes (r) sowie einer Reihe von Grafiken (tiff) gespeichert. Des Weiteren werden eine Readme-Datei (txt) sowie der Fragebogen (pdf/a) und ein Codebook (pdf/a) zur Beschreibung der Daten erstellt. Zusätzlich zu den selbst erhobenen Forschungsdaten werden öffentlich zugängliche Daten in Form von öffentlichen Statistiken (csv), Berichten (docx, pdf) und gesetzlichen Regelungen (html, pdf) für das Projekt wiederverwendet und referenziert. Die zu erwartende Gesamtgröße aller Dateien beträgt maximal 50 GB.

Dokumentation und Datenqualität: Metadaten werden über das Webformular des GESIS - Leibniz-Institut für Sozialwissenschaften nach dem fachspezifischen DDI-Standard erstellt. Eine zusätzliche Dokumentation der Forschungsdaten in Form von Readme, Fragebogen, Codebook und R-Syntax ist ebenso vorgesehen. Schlagwörter werden nach dem fachspezifischen Thesaurus TheSoz vergeben. Die Daten werden im Rahmen des Webformulars mittels der fachspezifischen Klassifikation Sozialwissenschaften klassifiziert. Die Qualität der Daten wird durch statistische Verfahren vor allem hinsichtlich der Repräsentativität und Reliabilität geprüft. Hierzu wird bspw. die Teilnahme der Personen mit dem entsprechenden Anteil der amtlichen Statistik verglichen und es erfolgt bei Bedarf eine Gewichtung. Für die Nutzung der erhobenen Daten ist eine Tabellenkalkulationssoftware, ein Textverarbeitungsprogramm, eine Statistiksoftware sowie ein PDF-Viewer notwendig.

Speicherung und technische Sicherung während des Projektverlaufs: Die sichere Speicherung und das Backup werden während der Projektlaufzeit durch die Projektleitung in Kooperation mit dem zuständigen DV-Beauftragten des Instituts XYZ sichergestellt. Für die Speicherung und kollaborative Verarbeitung der Daten während des Projekts wird der universitätseigene Cloudspeicher „HU-Box“ verwendet. Dies ermöglicht ein klares Zugriffsmanagement und eine einfache Nutzungsverwaltung. Für sensible Daten werden verschlüsselte, passwortgeschützte Ordner verwendet, die nur von berechtigten Mitarbeitenden eingesehen und verarbeitet werden können. Ein Backup erfolgt automatisiert jede Nacht.

Rechtliche Verpflichtungen und Rahmenbedingungen: Die Teilnehmenden werden im Rahmen des Onlinefragebogens über eine spätere Veröffentlichung unter Wahrung der Anonymität informiert. Die Online-Umfrage wird DSGVO-konform unter Beratung durch die institutionellen Datenschutzbeauftragten erstellt. Dies beinhaltet eine informierte Einwilligung der Befragten und die separate Einwilligung zur späteren Publikation der erhobenen Daten. Ein Ethikvotum wird zuvor bei der zuständigen Ethikkommission der Humboldt-Universität eingeholt. Zur urheberrechtlichen Klärung der Eigentumsrechte an den Daten wird ein Kooperationsvertrag mit Projektpartner Z geschlossen sowie ein Datenmanagementplan im Projekt erstellt.

Datenaustausch und dauerhafte Zugänglichkeit der Daten: Neben der direkten Auswertung durch die Projektgruppe wird der Datensatz auch für andere Forschungsprojekte von Relevanz sein. Da bisher keine vergleichbaren Daten für eine Sekundäranalyse zur Verfügung stehen, werden die Forschungsdaten, die erstellten R-Analyseskripte und der

¹ Die Inhalte dieser Vorlage orientieren sich an der Checkliste zum Umgang mit Forschungsdaten der Deutschen Forschungsgemeinschaft in der Version vom 21.12.2021. Die Leitfragen der DFG-Checkliste sind in den Kommentaren beigelegt.

Kommentiert [KH1]: Auf welche Weise entstehen in Ihrem Projekt neue Daten? Werden existierende Daten wiederverwendet? Welche Datentypen, im Sinne von Datenformaten (z. B. Bilddaten, Textdaten oder Messdaten) entstehen in Ihrem Projekt und auf welche Weise werden sie weiterverarbeitet? In welchem Umfang fallen diese an bzw. welches Datenvolumen ist zu erwarten?

Kommentiert [KH2]: Welche Ansätze werden verfolgt, um die Daten nachvollziehbar zu beschreiben (z. B. Nutzung vorhandener Metadaten- bzw. Dokumentationsstandards oder Ontologien)? Welche Maßnahmen werden getroffen, um eine hohe Qualität der Daten zu gewährleisten? Sind Qualitätskontrollen vorgesehen und wenn ja, auf welche Weise? Welche digitalen Methoden und Werkzeuge (z. B. Software) sind zur Nutzung der Daten erforderlich?

Kommentiert [KH3]: Auf welche Weise werden die Daten während der Projektlaufzeit gespeichert und gesichert? Wie wird die Sicherheit sensibler Daten während der Projektlaufzeit gewährleistet (Zugriffs- und Nutzungsverwaltung)?

Kommentiert [KH4]: Welche rechtlichen Besonderheiten bestehen im Zusammenhang mit dem Umgang mit Forschungsdaten in Ihrem Projekt? Sind Auswirkungen oder Einschränkungen in Bezug auf die spätere Veröffentlichung bzw. Zugänglichkeit zu erwarten? Auf welche Weise werden nutzungs- und urheberrechtliche Aspekte sowie Eigentumsfragen berücksichtigt? Existieren wichtige wissenschaftliche Kodizes bzw. fachliche Normen, die Berücksichtigung finden sollten?

Kommentiert [KH5]: Welche Daten bieten sich für die Nachnutzung in anderen Kontexten besonders an? Nach welchen Kriterien werden Forschungsdaten ausgewählt, um diese für die Nachnutzung durch andere zur Verfügung zu stellen? Planen Sie die Archivierung Ihrer Daten in einer geeigneten Infrastruktur? Falls ja, wie und wo? Gibt es Sperrfristen? Wann sind die Forschungsdaten für Dritte nutzbar?

zugehörige Fragebogen für andere Forschende im GESIS - Leibniz-Institut für Sozialwissenschaften unter einer CC-BY Lizenz zur Verfügung gestellt. Über das GESIS-Datenarchiv erhält die Studie einen Digital Object Identifier (DOI). Wie in den Leitlinien zur guten wissenschaftlichen Praxis der DFG vorgesehen, werden die Projektergebnisse und alle relevanten Forschungsdaten darüber hinaus für mindestens zehn Jahre an der Humboldt-Universität zu Berlin gespeichert. Hierzu wird mit dem institutionellen DV-Beauftragten ein Verfahren zur Überführung der Daten in den Backup-Service des CMS vereinbart. Die Kuration der Daten nach Projektende erfolgt durch das GESIS - Leibniz-Institut für Sozialwissenschaften.

Verantwortlichkeiten und Ressourcen: Entsprechend der Grundsätze zum Umgang mit Forschungsdaten an der Humboldt-Universität zu Berlin (<https://hu.berlin/forschungsdaten-policy>) ist die Projektleitung für alle Aspekte des Forschungsdatenmanagements verantwortlich. Einzelne Teilbereiche sollen jedoch an die Projektmitarbeiter:innen delegiert werden. So sind für die Aufbereitung der Forschungsdaten für die Veröffentlichung im Repozitorium 3 PM vorgesehen. Die Verfügbarmachung und Archivierung der Daten durch das GESIS - Leibniz-Institut für Sozialwissenschaften ist nach Rücksprache mit dem Repozitorium kostenfrei. Die langfristige Speicherung der Daten erfolgt kostenfrei über den CMS der HU Berlin.

Kommentiert [KH6]: Wer ist verantwortlich für den adäquaten Umgang mit den Forschungsdaten (Beschreibung der Rollen und Verantwortlichkeiten innerhalb des Projekts)? Welche Ressourcen (Kosten; Zeit oder anderes) sind erforderlich, um einen adäquaten Umgang mit Forschungsdaten im Projekt umzusetzen? Wer ist nach Ende der Laufzeit des Projekts für das Kuratieren der Daten verantwortlich?



Zitationsvorschlag: FDM-Team der HU (2021):
Konzept zum Umgang mit Forschungsdaten.
Musterplan auf Grundlage der DFG-Checkliste.