

2023

DATEN- & METHODENBERICHT, SUF VERSION 1.0



**FORSCHUNGSZENTRUM
DEMOGRAFISCHER
WANDEL (FZDW)**

**ENTWICKLUNG UND VERBREITUNG VON FORSCHUNGSDATEN-
MANAGEMENT AN FACHHOCHSCHULEN UND HOCHSCHULEN
FÜR ANGEWANDTE WISSENSCHAFTEN (EVER_FDM)**
DATEN- & METHODENBERICHT
ROBERT WERTH



Autor:

Robert Werth 

Zitationsempfehlung:

Werth, R. (2023). *Daten- und Methodenbericht: Entwicklung und Verbreitung von Forschungsdatenmanagement an Fachhochschulen und Hochschulen für Angewandte Wissenschaften, Scientific Use File Version 1.0*. Frankfurt am Main, Germany, Forschungszentrum Demografischer Wandel (FZDW), Frankfurt University of Applied Sciences.

Die Durchführung der EVER_FDM-Studie war eine große Gemeinschaftsleistung. Der Beitrag der folgenden Personen wird dabei dankend anerkannt:

Prof. Dr. Andreas Klocke
Dr. Sven Stadtmüller
Arnela Balic
Meike Backes
Lukas Magnin

Kontakt:

Frankfurt University of Applied Sciences
Forschungszentrum Demografischer Wandel (FZDW)
Nibelungenplatz 1
60318 Frankfurt am Main
E-Mail: info@fzdw.de
Web: www.fzdw.de



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

EVER_FDM wurde:



Förderkennzeichen: 16DFH201

Inhaltsverzeichnis

1. Übersicht zur Erhebung.....	3
2. Datennutzungshinweise	5
3. Inhalte und Anlage der Befragung.....	7
4. Methodisches Vorgehen	9
5. Datenaufbereitung	11
5.1 Datenübertragung	11
5.2 Codierung offener Angaben	11
5.3 Datenprüfung und Datenbereinigung	11
5.4 Generierung von Variablen	12
5.5 Datenstruktur und Dateiformat	12
5.6 Variablennamen, Labels und Missings	12
6. Anonymisierung.....	14
7. Literaturverzeichnis.....	16

1. Übersicht zur Erhebung

Studie	Entwicklung und Verbreitung von Forschungsdatenmanagement an Fachhochschulen und Hochschulen für Angewandte Wissenschaften
Akronym	EVER_FDM
Erhebende Institution	Frankfurt University of Applied Sciences (Frankfurt UAS)
Gefördert vom	Bundesministerium für Bildung und Forschung (BMBF)
Förderkennzeichen	16DFDH201
Projektlaufzeit	01.06.2022 bis 31.05.2023
Feldzeit	03.11.2022 bis 31.01.2023
Projektleitung	Prof. Dr. Andreas Klocke, ORCID: 0000-0001-5842-4875
Mitarbeiter(in) #1	Robert Werth, ORCID: 0000-0003-0215-2545
Mitarbeiter(in) #2	Arnela Balic, ORCID: 0000-0002-7049-5737
Projektbeschreibung	Die Erhebung wurde als standardisierte Online-Befragung durchgeführt, die vor allem drei Themenbereiche in Bezug auf Forschungsdatenmanagement (FDM) umfasste: FDM-Awareness, FDM-Aktivitäten und FDM-Bedarfe. Es handelte sich um eine Vollerhebung an allen staatlichen FHs/HAWen in Deutschland und die Zielgruppen waren Professor*innen sowie wissenschaftliche Mitarbeiter*innen an diesen Einrichtungen. Zunächst wurde an jeder Hochschule sowohl eine Person aus dem Präsidium/Rektorat als auch eine für das für Forschungsdatenmanagement (FDM) zuständige Person identifiziert und per E-Mail kontaktiert. Diese Personen wurden gebeten, den Link zum Fragebogen an das wissenschaftliche Personal ihrer Hochschule weiterzuleiten. Auf diese Weise konnten 2.456 Personen von 109 Hochschulen für die Teilnahme gewonnen werden. Ein Pretest sicherte die Qualität des Fragebogens. Die Aufbereitung und Auswertung der Daten erfolgte syntaxbasiert mit Stata.
Schlagwörter	Forschungsdatenmanagement, Forschungsdateninfrastruktur, Gute Wissenschaftliche Praxis, Forschungsdokumentation, FAIR, Wissenschaftsplanung, Datenverwaltung, Wissenschaftliches Arbeiten, Datendokumentation, Wissenschaftsmanagement, Forschungsdesign
Erhebungsdesign	Querschnitt
Erhebungsdatentyp	Quantitative Daten
Erhebungsmethode	Standardisierte Online-Befragung
Grundgesamtheit	Forschende an Hochschulen für Angewandte Wissenschaften in Deutschland
Stichprobe	Vollerhebung
Teilnehmende Hochschulen	109 von 210 eingeladenen Hochschulen für Angewandte Wissenschaften (51,9 %)
Nettostichprobe	n = 2.456 (Teilnehmende) n = 2.154 (Forschende)

Rücklaufquote	Unbekannt, da keine eindeutige Klassifikation als „forschend“ vorhanden.
Datenprodukt und Zugangsweg	SUF: Download
Datensatzstruktur	Personendatensätze im wide-Format
DOI	https://doi.org/10.21249/DZHW:everfdm:1.0.0
Zitation	Klocke, A., Werth, R. & Balic, A. (2023). Entwicklung und Verbreitung von Forschungsdatenmanagement an Fachhochschulen und Hochschulen für Angewandte Wissenschaften (EVER_FDM). Scientific Use File Version 1.0.0. Frankfurt am Main: Forschungszentrum Demografischer Wandel (FZDW). https://doi.org/10.21249/DZHW:everfdm:1.0.0
Anmerkungen	Die Daten wurden vom Forschungszentrum Demografischer Wandel (FZDW) der Frankfurt UAS im Rahmen einer Förderung durch das Bundesministerium für Bildung und Forschung (BMBF) erhoben. Neben den veröffentlichten Daten aus der Online-Befragung wurden im Projekt zudem Daten aus fünf qualitativen, leitfadengestützten Gruppendiskussionsrunden erhoben. Dabei wurden FDM-Sachkundige aus Verwaltung und Forschungsinfrastruktur zu ihren Erfahrungen, Bedarfen, Problemen und Lösungen im Hinblick auf die Etablierung von FDM interviewt. Es handelte sich also um eine mixed-methods Studie. Aus datenschutzrechtlichen Gründen werden die Daten der qualitativen Erhebung nicht zur Nachnutzung bereitgestellt.
Weitere Informationen	https://fzdw.de/projekte/ever_fdm/ https://forschungsdaten.info/fdm-im-deutschsprachigen-raum/deutschland/hessen/ever-fdm/

2. Datennutzungshinweise

[Voraussetzungen der Datennutzung] Die Daten der EVER_FDM Online-Erhebung werden durch das FDZ des DZHW entsprechend der europäischen Datenschutzgrundverordnung (EU-DSGVO) anonymisiert bereitgestellt und ausschließlich zur wissenschaftlichen Nutzung freigegeben.¹ Das FDZ bietet ein Scientific Use File (SUF) für die wissenschaftliche Sekundärnutzung an.

Voraussetzungen für die Nutzung des SUF sind die Anstellung der Datennutzerin/des Datennutzers an einer wissenschaftlichen Einrichtung und der Abschluss eines Datennutzungsvertrags mit dem FDZ. Studierende oder Promovierende ohne eine Anstellung an einer wissenschaftlichen Einrichtung müssen gemeinsam mit einer/einem betreuenden Mitarbeiter*in einen Datennutzungsvertrag abschließen. Im Zuge des Vertragsabschlusses wird durch das FDZ auch das Vorliegen eines wissenschaftlichen Nutzungsinteresses geprüft. Das Formular für den Datennutzungsantrag kann von der Website des FDZ heruntergeladen werden.

[Datenzugang] Die Daten werden verschlüsselt auf der Website des FDZ zum **Download** bereitgestellt. Datennutzende können die Daten auf ihrem lokalen Computer speichern, falls gewünscht selbst mit Daten aus externen Quellen verknüpfen und die Daten mit eigener Software analysieren.

[Datenprodukte] Über den Digital Object Identifier (DOI) 10.21249/DZHW:everfdm:1.0.0 ist eine Website mit zentralen Informationen zur Studie, weiteren Dokumentationsmaterialien sowie einer Übersicht der zur Verfügung stehenden Datenprodukte zur Studie erreichbar.

[Gebühren der Datenbereitstellung] Das SUF wird derzeit (Stand: November 2023) kostenfrei zur Verfügung gestellt. Änderungen bzw. die aktuelle Gebührenordnung können auf der Website des FDZ (<https://fdz.dzhw.eu>) eingesehen werden.

[Pflichten der Datennutzenden] Die Datennutzenden sind verpflichtet, folgende Regeln² einzuhalten:

- **Wissenschaftliche Nutzung:** Die Daten dürfen ausschließlich für wissenschaftliche Zwecke verwendet werden. Eine kommerzielle Nutzung ist untersagt.
- **De-Anonymisierungsverbot:** Jeder Versuch der Re-Identifikation von Analyseeinheiten (z. B. Personen, Haushalten, Institutionen) ist verboten.
- **Gebot zur Mitteilung von Sicherheitslücken:** Falls Datennutzende Kenntnis von Sicherheitslücken hinsichtlich Datenschutz bzw. Datensicherheit erlangen, müssen diese dem FDZ unverzüglich angezeigt werden.
- **Keine Weitergabe der Daten:** Das SUF darf nur durch die Person genutzt werden, die den Datennutzungsvertrag abgeschlossen hat.

¹ Das Datenschutzkonzept des FDZ ist angelehnt an den Portfolio-Ansatz von Lane, Heus und Mulcahy (2008, S. 6 ff.), an dem sich bereits das Leibniz-Institut für Bildungsverläufe (LifBi) (Koberg, 2016, S. 699 ff.) und das FDZ der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (Hochfellner, Müller, Schmucker und Roß, 2012, S. 9 f.) orientieren. Das FDZ des DZHW hat diesen Ansatz an die Anforderungen der eigenen Datenbestände angepasst und nutzt vier Kategorien von Maßnahmen zur Sicherstellung des Datenschutzes, die in unterschiedlicher Weise kombiniert werden können: Rechtlich-institutionelle Maßnahmen, informationelle Maßnahmen, technische Maßnahmen und statistische Maßnahmen.

² Der Datennutzungsvertrag regelt die Nutzungsbedingungen im Detail.

- **Löschungsgebot:** Das SUF ist nach Ablauf der vereinbarten Nutzungsdauer (in der Regel 1,5 Jahre) von jeglichen Rechnern, Servern und Datenträgern zu löschen. Ebenso müssen alle Sicherungskopien, modifizierten Datensätze (z. B. Arbeits-, Auszugs- oder Hilfsdateien) sowie Ausdrücke vernichtet werden.
- **Bereitstellung/Meldung von Publikationen:** Jede Art von Publikation, die aus der Arbeit mit Daten des FDZ hervorgeht, ist dem FDZ unmittelbar nach Veröffentlichung anzuzeigen. Dabei ist dem FDZ eine elektronische Version der Druckfassung zur Verfügung zu stellen.
- **Zitationspflicht:** Die verwendeten Daten müssen in Veröffentlichungen, anderen Arbeiten (z. B. Abschlussarbeiten) und Vorträgen laut der Vorgaben des FDZ zitiert werden.

3. Inhalte und Anlage der Befragung

[Ziele] Mit der Online-Befragung im Rahmen der EVER_FDM Studie wurde erstmals eine deutschlandweite Datenbasis geschaffen, mit der die Situation der Fachhochschulen und Hochschulen für Angewandte Wissenschaften (HAWen) hinsichtlich des Forschungsdatenmanagements (FDM) verlässlich eingeschätzt werden kann. Die Untersuchung dient dazu, nationale Maßnahmen zur Förderung von FDM (z. B. im Bereich Schulung, Beratung oder technischer Infrastruktur) passgenau für diesen Hochschultypus (weiter-)entwickeln zu können.

Zudem profitierten auch die beteiligten Hochschulen in besonderer Weise von den Ergebnissen, da sie hochschulspezifische Ergebnisberichte anfordern konnten. So bekamen sie einerseits einen Überblick über die Aktivitäten und Bedarfe der Forschenden an ihrer Hochschule und konnten sich andererseits mit dem bundesdeutschen Durchschnitt vergleichen. Sie gelangten so zu einer besseren Einschätzung der Problemlagen und des Wissensstands zum FDM bei ihren Forschenden und konnten Maßnahmen der FDM-Unterstützung strategisch planen.

[Themenfelder] In der Online-Erhebung der EVER_FDM-Studie wurden vor allem drei Themenbereiche in den Blick genommen:

- FDM-Awareness
- FDM-Aktivitäten
- FDM-Bedarfe

Bei der FDM-Awareness wurde erhoben, inwieweit FDM als Thema unter den Forschenden überhaupt bekannt ist. So konnte eine Einschätzung gelingen, wie relevant FDM bereits im Forschungsalltag ist und ob womöglich zunächst eine grundlegende Sensibilisierung erreicht werden muss. Hierbei wurde bspw. erfragt, welche FDM-Richtlinien (bspw. die FAIR-Prinzipien) und FDM-Institutionen (bspw. die NFDI) bei den Forschenden bereits bekannt waren.

Die Abfrage der FDM-Aktivitäten bezog sich darauf, welche Maßnahmen des FDM die Forschenden bereits umsetzten (z. B. ob sie bereits einmal Daten veröffentlicht oder auf andere Weise anderen Forschenden zur Verfügung gestellt haben). Dies ist potenziell auch ohne FDM-Awareness möglich, denn alle Forschenden müssen zwangsläufig ihre Daten in irgendeiner Weise organisieren. Das bedeutet, Forschende betreiben womöglich bereits (professionelles) FDM, ohne sich dessen bewusst zu sein.

Der letzte und zugleich wichtigste Teil der Erhebung betraf die FDM-Bedarfe. Hierbei wurde erfasst, was sich Forschende im Bereich FDM wünschen. Dabei galt es den Zwiespalt zwischen einer offenen oder geschlossenen Abfrage der Bedarfe zu überwinden. Werden Bedarfe offen abgefragt, ist mit erhöhter Item-Nonresponse zu rechnen, da den Befragten entweder nichts Konkretes einfällt, oder sie schlichtweg Zeit sparen möchten und diese Fragen überspringen. Stellt man hingegen geschlossene Fragen, werden die Bedarfe womöglich erst durch die Befragung selbst geweckt. Um diesem Problem zu begegnen, setzte die Befragung auf eine Kombination der beiden Strategien. Zunächst wurden spontane Bedarfe und Assoziationen offen abgefragt und dann anhand geschlossener Fragen konkrete Themenbereiche behandelt. Diese betrafen sowohl die technische Infrastruktur (z. B. Speicherplatz, Cloud-Dienste, Datenmanagementplansoftware, Elektronische Laborbücher, Umfragesoftware, Repositorien usw.) als auch Bedarfe hinsichtlich Weiterbildung, Beratung, Vernetzung oder verbindlicher Policies und Richtlinien.

Um differenzierte Analysen zu ermöglichen, wurden zudem einige Sozialstrukturmerkmale der Befragten erfasst (z. B. Alter, Geschlecht, Fachrichtung und Statusgruppe).

[Qualitätssicherung und Pretest] Um die Verständlichkeit, Vollständigkeit, Angemessenheit und methodische Qualität des Erhebungsinstruments sowie einen geringen response burden (d. h. Aufwand, der nötig ist, um den Fragebogen auszufüllen) sicherzustellen, wurden verschiedene Maßnahmen ergriffen:

- Zunächst wurden bestehende FDM-Umfragen gesichtet, um geeignete Fragen für die Nachnutzung zu identifizieren und ggf. eine Vergleichbarkeit zu gewährleisten. Dabei wurden einige Fragen aus anderen Erhebungen angepasst oder vollständig übernommen (z. B. Fachhochschule Potsdam 2018, Technische Universität Hamburg 2016, Christian-Albrechts-Universität zu Kiel 2014). Zudem wurde für die Abfrage der Fachrichtung, die Fächersystematik des Statistischen Bundesamts verwendet (Statistisches Bundesamt, 2023).
- Nach initialer Fertigstellung des Fragebogens wurde dieser verschiedenen FDM-Sachkundigen zur Prüfung auf Vollständigkeit und inhaltliche Passung vorgelegt. Besonders hervorzuheben sind dabei die Beiträge des BMBF-Projekts FDMH-Nord sowie der Referentinnen und Referenten für Forschungsdatenmanagement der Landesinitiative Hessische Forschungsdateninfrastrukturen (HeFDI).
- Anschließend wurde der Fragebogen von Sven Stadtmüller vom Team Survey Design and Methodology des GESIS Leibniz-Instituts für Sozialwissenschaften auf methodische Qualität überprüft.
- Die adäquate Organisation und Durchführung der Befragung wurde im Prozess der Verfahrensdokumentation durch die Kompetenzstelle IT-Dokumentation der Frankfurt UAS sichergestellt. Hierbei wurde die gesamte Planung des Datenmanagements begutachtet und alle Unterlagen insb. hinsichtlich datenschutzrechtlicher Aspekte geprüft. Anschließend erfolgte eine Freigabe durch den Datenschutzbeauftragten sowie den Kanzler der Frankfurt UAS.
- Der letzte Schritt im Qualitätssicherungsprozess bestand aus einem Pretest des Fragebogens. Dabei wurde eine Testversion des Fragebogens erstellt, welche sowohl mit offenen als auch geschlossenen Fragen zur Einschätzung von Einzelfragen sowie des gesamten Fragebogens angereichert war. Insbesondere sollten anhand einer vierstufigen Skala folgende Aussagen bewertet werden:
 1. „Die Fragen waren verständlich.“
 2. „Der Fragebogen war interessant.“
 3. „Der Fragebogen ist zu lang.“
 4. „Es hat Spaß gemacht den Fragebogen auszufüllen.“
 5. „Ich glaube, die Ergebnisse der Befragung können helfen, das Forschungsdatenmanagement an meiner Hochschule zu verbessern.“

Anschließend wurden aus jedem der vier Fachbereiche der Frankfurt UAS drei Professor*innen und zwei wissenschaftliche Mitarbeiter*innen zur Teilnahme am Pretest eingeladen (N=20). Neun der eingeladenen Personen kamen der Aufforderung nach (n=9).

- Vor Veröffentlichung erfolgte zudem eine technische, inhaltliche und rechtliche Kuration der Daten durch das Forschungsdatenzentrum für Hochschul- und Wissenschaftsforschung (FDZ-DZHW).

4. Methodisches Vorgehen

[Grundgesamtheit] Die Grundgesamtheit der Erhebung bilden alle forschend tätigen Personen an deutschen Fachhochschulen bzw. Hochschulen für Angewandte Wissenschaften. Dies umfasst sowohl die öffentlich-rechtlichen als auch die privaten und kirchlichen HAWen, sofern diese staatlich anerkannt sind. Zudem sind auch die Hochschulen eigenen Typs Teil der Grundgesamtheit. Universitäten, Verwaltungshochschulen sowie künstlerische Hochschulen sind nicht Teil der Untersuchung.

Innerhalb dieser Hochschulen (zum Erhebungszeitpunkt 210 Stück) war ausschließlich das Personal Adressat der Befragung, welches in der Forschung tätig ist oder war. Die genaue Größe dieser Grundgesamtheit kann nicht eindeutig bestimmt werden. Zwar waren im Jahr 2020 insgesamt 102.097 Personen als wissenschaftliches oder künstlerisches Personal an deutschen HAWen beschäftigt (Statistisches Bundesamt, 2021), es kann jedoch nicht davon ausgegangen werden, dass all diese Personen auch forschend tätig sind oder es waren. Dies liegt darin begründet, dass an HAWen nach wie vor die Lehre im Fokus steht (Bad Wiesseer Kreis, 2019) und das Lehrdeputat der Professor*innen mit 18 Semesterwochenstunden doppelt so hoch ist wie an Universitäten. Zudem haben HAWen nur in wenigen Bundesländern (z. B. Hessen) ein eigenständiges Promotionsrecht, wodurch ein gleichwertiger akademischer Mittelbau oft fehlt. Um die Eingrenzung der Befragten auf die Grundgesamtheit der Forschenden sicherzustellen, beinhaltete der Fragebogen Kontrollfragen zur Statusgruppe und zur Forschungsaktivität der Teilnehmenden.

[Stichprobe] Die Untersuchung war als Vollerhebung ausgelegt, d. h. alle 210 staatlichen bzw. staatlich anerkannten HAWen waren Teil des Samples und es sollten alle forschend tätigen Beschäftigten zur Teilnahme an der Befragung eingeladen werden.

[Durchführung der Erhebung] Zunächst wurden per Internetrecherche von jeder HAW ein zuständiges Mitglied der Hochschulleitung (z. B. Vizepräsident*in für Forschung) sowie eine zuständige Person für das Forschungsdatenmanagement (z. B. FDM Scout, Forschungsreferent*in etc.) identifiziert. Konnte keine dezidierte Person gefunden werden, wurden die jeweiligen allgemeinen Kontaktdaten der Hochschule bzw. der Forschungsabteilung notiert.

Am 03.11.2022 wurden alle Kontaktpersonen per E-Mail kontaktiert und gebeten den Teilnahmelink an das wissenschaftliche Personal ihrer Hochschule weiterzuleiten. Die Einladung beinhaltete zudem einen Link auf ein Ansichtsexemplar des Fragebogens (im PDF-Format) sowie einen Textvorschlag für die E-Mail an die Forschenden. Außerdem wurde auf die Möglichkeit der Ausfertigung eines hochschulspezifischen Ergebnisberichts hingewiesen.

Am 15.12.2022 wurde ein Erinnerungsschreiben per E-Mail an die Kontaktpersonen versandt. Dabei wurde auch die bisherige Anzahl der Teilnehmenden der jeweiligen Hochschule kommuniziert. Am 31.01.2023 wurde die Umfrage geschlossen, was einer Feldzeit von knapp drei Monaten entspricht.

[Rücklaufsteigernde Maßnahmen] Um die Akzeptanz der Befragung zu steigern und dadurch den Rücklauf zu erhöhen, wurden im Vorfeld der Befragung verschiedene Maßnahmen ergriffen. So wurde zum Projektstart eine Pressemitteilung verfasst und eine Informationsseite zur Studie auf der Webseite des FZDW eingerichtet (https://fzdw.de/projekte/ever_fdm/). Zudem wurden regelmäßige Projektupdates auf der Seite „Aktuelles“ des FZDW verfasst (<https://fzdw.de/aktuelles/>) und Projekt-

beschreibungen auf den FDM-Seiten [Forschungsdaten.info](https://www.forschungsdaten.info) und [Forschungsdaten.org](https://www.forschungsdaten.org) platziert. Zudem wurden das Projekt und insb. die Online-Erhebung auf verschiedenen Tagungen und Veranstaltungen vorgestellt:

- AG "FDM-Strategien für FH/HAW" der Landesinitiative fdm.nrw am 31.08.2022
- Fortrama Jahrestagung am 12.10.2022
- Live-Treffen der Landesinitiative Hessische Forschungsdateninfrastrukturen (HeFDI) am 17.10.2022
- Bundesweite Vernetzungsveranstaltung FDM an HAW am 18.10.2022

Darüber hinaus fand eine enge Zusammenarbeit mit dem BMBF-Projekt FDMH-Nord statt, welches die Befragung an den HAWen in Schleswig-Holstein verbreitete. Das Angebot, nach Projektabschluss einen hochschulspezifischen Ergebnisbericht anfordern zu können, welches allen teilnehmenden Hochschulen mit einer Teilnehmendenzahl von über 15 Forschenden gemacht wurde, kann ebenfalls zu den rücklaufsteigernden Maßnahmen gezählt werden.

[Rücklauf] Insgesamt konnten von 109 der 210 eingeladenen HAWen Daten gewonnen werden. Dies entspricht einer Teilnahmequote von 52 %. Insgesamt nahmen 2.456 Personen an der Befragung teil. Davon gaben 2.154 an, zum Befragungszeitpunkt oder vorher an der aktuellen Hochschule geforscht zu haben. Die Berechnung einer dezidierten Rücklaufquote ist aufgrund der unter „[Grundgesamtheit]“ beschriebenen Sachverhalte leider nicht möglich. Betrachtet man die Gesamtzahl der Teilnehmenden, so ist EVER_FDM die zum Durchführungszeitpunkt umfangreichste Datenerhebung zum Forschungsdatenmanagement in Deutschland.

5. Datenaufbereitung

Im Folgenden werden die verschiedenen Schritte der Datenübertragung und -aufbereitung beschrieben. Die Online-Erhebung wurde mit der Software Questionstar durchgeführt. Die weitere Verarbeitung fand syntaxbasiert mit der Software Stata (Version 14) statt.

5.1 Datenübertragung

Die Angaben wurden zunächst als SPSS-Datei (.sav) aus der Befragungssoftware exportiert. Dadurch waren bestimmte Metadaten (insb. Variablen- und Wertelabels) bereits im Datensatz enthalten. Anschließend wurde die Datei in das Stata-Format (.dta) konvertiert und lokal weiterverarbeitet.

5.2 Codierung offener Angaben

Im Rahmen der Datenaufbereitung erfolgte eine Codierung der (halb-)offenen Angaben im Fragebogen. Dafür wurden zunächst Codierlisten erstellt, welche zusammenfassende Kategorien möglicher Antworten enthalten. Diese basierten auf einer überblicksartigen Sichtung der gegebenen Antworten. Auf bereits bestehende Codierschemata konnte nicht zurückgegriffen werden, da diese für die EVER_FDM gestellten offenen Fragen nicht existierten. Aus Datenschutzgründen enthält das SUF lediglich die codierten Angaben.

Sofern offene Antworten einer Variablen mehreren Kategorien auf einmal zugeordnet werden konnten, wurde für jede Kategorie eine eigene neue Variable erstellt und diese dichotom mit „trifft zu“ oder „trifft nicht zu“ codiert. Ließen sich alle Angaben einer Variablen jeweils genau einer Kategorie zuordnen, so wurde nur eine codierte Variable mit verschiedenen Ausprägungen generiert. Bei einigen Variablen fand lediglich eine Recodierung statt. Das heißt, die offenen Angaben in der Residualkategorie einer Variablen wurden geprüft, ob diese sich nicht doch einer der geschlossen abgefragten Kategorien zuordnen lassen. Um die Vergleichbarkeit mit den Ursprungsangaben sicherzustellen, wurden für die Recodierungen neue Variablen gebildet.

Für Angaben, die sich nicht eindeutig einer Kategorie zuordnen ließen, wurden Codierkonventionen festgelegt. Diese sind im FZDW Technical Report „Generierte Variablen im EVER_FDM Datensatz“ (Werth, 2023) entsprechend dokumentiert.

5.3 Datenprüfung und Datenbereinigung

[Konsistenz- und Plausibilitätsprüfung] Im Anschluss an die Übertragung erfolgte eine umfassende Prüfung der Daten. Diese wurde syntaxbasiert mit Stata-Do-Files durchgeführt. Sie beinhaltete insb. die folgenden Arten von Prüfungen:

- *Prüfung der Einhaltung der Filterführung:* Es wurde überprüft, ob an den Stellen im Datensatz, an denen Fragen bestimmten Personen nicht gestellt (= überfiltert) werden sollten, auch tatsächlich ausschließlich fehlende Werte in diesen Fragen vorhanden waren.
- *Prüfung von Wertebereichen:* Es wurde überprüft, ob die erfassten Ausprägungen einer Variablen in dem für sie definierten Wertebereich lag.

[Löschung von Fällen] Alle Fälle, die ab Frage 3 bzw. Variable ac_q03_01 ausschließlich fehlende Werte enthielten, wurden aus dem Datensatz entfernt. Die Einhaltung der Zielgruppe wurde durch

die Fragen 1 und 2 (Variablen sd_q01 und sd_q02) geprüft, die Daten der Befragten ohne Forschungsaktivitäten wurden aus Gründen der Vollständigkeit jedoch im Datensatz belassen. Diese Personen wurden direkt zu Frage 23 (Variable aw_q23_01) weitergeleitet und beantworteten lediglich die Fragen zu FDM-Awareness und machten ggf. soziodemografische Angaben.

[Umkodierung] Die Ausprägungen dichotomer Variablen wurden einheitlich auf die Werte „0“ oder „1“ umkodiert.

5.4 Generierung von Variablen

Neben den Variablen mit den codierten Antworten der Befragten beinhaltet der EVER_FDM Datensatz auch weitere generierte Variablen. Dabei handelt es sich bspw. um Variablen, in denen Angaben aus Datenschutzgründen verändert (= anonymisiert) werden mussten. Des Weiteren wurden Hilfs- und Servicevariablen erstellt, welche die Arbeit mit den Daten erleichtern sollen. Hierzu zählen bspw. Kategorisierungen oder Variablen, die Informationen aus mehreren anderen Variablen enthalten. Die Entstehungslogiken der generierten Variablen inkl. der zugehörigen Stata Syntaxen sind im FZDW Technical Report „Generierte Variablen im EVER_FDM Datensatz“ (Werth, 2023) detailliert beschrieben.

5.5 Datenstruktur und Dateiformat

Der Datensatz enthält die Befragungsdaten sowie die zusätzlich generierten Variablen und Daten, welche die Fragebogensoftware automatisiert generiert hat (bspw. die Befragungsdauer). Die Reihenfolge der Variablen orientiert sich an der Reihenfolge der zugehörigen Fragen im Fragebogen. Die generierten Variablen wurden jeweils nach den Ursprungsvariablen einsortiert. Der Datensatz wird sowohl im Stata- als auch im SPSS-Format bereitgestellt.

5.6 Variablennamen, Labels und Missings

[Variablenbenennung] Für die Benennung der Variablen wurde ein einheitliches Benennungsschema verwendet. Die durch einen Unterstrich getrennten Elemente stehen dabei für verschiedene Charakteristika, die den Nutzenden Orientierung bzgl. der Inhalte der Variablen geben. Der Aufbau ist standardisiert, was eine automatisierte Weiterverarbeitung erleichtert. Das Präfix der Variablen enthält das jeweilige Themenfeld. Anschließend folgt nach einem „q“ (= question) die Nummer der Frage im Fragebogen. Anschließend folgt einer oder mehrere Suffixe, die Zusatzinformationen enthalten, wie die Kenntlichmachung von Variablen mit offenen Angaben oder von generierten Variablen. Die enthaltenen Elemente stehen für die folgenden Inhalte.

Präfix:

sy_ = Prozessgenerierte Variablen (system)

aw_ = Awareness (awareness)

ac_ = Aktivitäten (activities)

ne_ = Bedarfe (needs)

sd_ = Soziodemographie (sociodemography)

Variablenname:

q01, q02, ... , q## = Laufende Nummer im Fragebogen

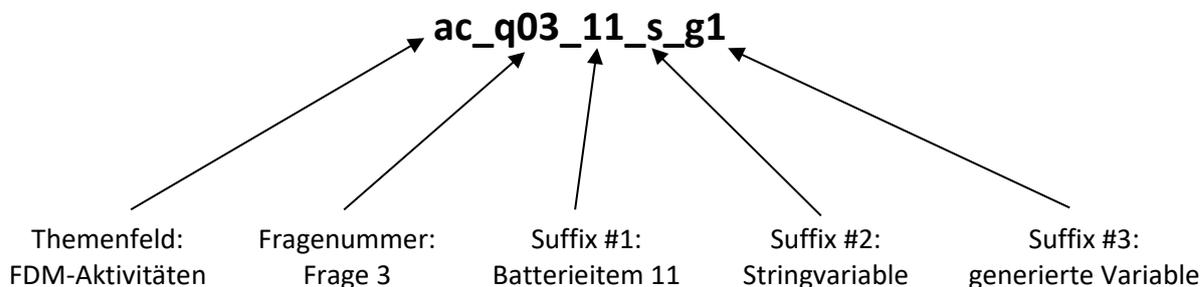
Suffixe:

`_01, _02, ... , _##` = Wird verwendet, wenn die Variable Teil einer Itembatterie ist

`_s` = Zeigt an, dass es sich um eine String-Variable (offene Texteingabe) handelt

`_g1, _g2, ... , _g#` = Zeigt an, dass es sich um eine generierte Variable handelt

Abbildung 1: Beispiel des Benennungsschemas der EVER_FDM Online-Erhebung



[Variablen- und Wertelabels] Für Wertelabels wurden die Bezeichnungen der Werte im Fragebogen übernommen. Für Variablenlabels wurden Kurzformen der Formulierungen im Fragebogen gewählt. Bei Fragen mit Mehrfachantwort sowie Itembatterien wurde eine Kurzform der Ursprungsfrage verwendet und dann nach einem Doppelpunkt die jeweilige Antwortoption bzw. die jeweilige Itemformulierung als Variablenlabel übernommen. Codierte Variablen wurden zusätzlich zur Kennzeichnung mit dem Präfix `_g#` im Variablennamen, mit dem Hinweis „(codiert)“ oder „(recodiert)“ im Variablenlabel versehen.

[Codierung fehlender Werte] Zur Codierung fehlender Werte wurde eine einheitliche Missing-Systematik verwendet. Fehlende Angaben werden dabei mit negativen Werten kenntlich gemacht. Die verwendete Systematik entspricht dabei den Empfehlungen des FDZ DZHW für die Codierung fehlender Werte (Verbund Forschungsdaten Bildung, 2019).

Tabelle 1: Systematik fehlender Werte in der EVER_FDM Online-Erhebung

Wert	Wertelabel	Beschreibung
-989	filterbedingt fehlend	Wenn eine Frage im Fragebogen überfiltert wurde.
-966	nicht ermittelbar	Wenn bei der Generierung einer Variablen kein Wert vergeben werden konnte, weil nicht genug Ausgangsinformationen vorlagen.
-968	unplausibler Wert	Wenn bei einer offenen Frage keine inhaltlichen Angaben gemacht wurden (z. B. nur Satzzeichen wie „-“, „/“ oder „.“).
-999	weiß nicht	Wenn Option im Fragebogen ausgewählt wurde.
-949	kann ich nicht einschätzen	Wenn Option im Fragebogen ausgewählt wurde.
-950	Ich übe keine Lehrtätigkeit aus.	Wenn Option im Fragebogen ausgewählt wurde.
-997	keine Angabe (Antwortkategorie)	Wenn Option im Fragebogen ausgewählt wurde.
-998	keine Angabe	Wenn anderen Gründe für das Fehlen des Wertes vorlagen.

6. Anonymisierung

In der EVER_FDM Online-Erhebung wurden weder Namen noch Kontaktdaten irgendwelcher Art von den Teilnehmenden erhoben. Auch die Ansprache erfolgte anonym, da die Kontaktpersonen an den Hochschulen als Gatekeeper fungierten, die den Fragebogen jeweils an einen dem FZDW nicht bekannten Adressatenkreis weiterleiteten. Insofern handelte es sich um eine formal anonyme Erhebung.

Abhängig vom Antwortverhalten (insb. bei der Beantwortung der offenen Fragen) war jedoch nicht vollständig auszuschließen, dass die Befragten Angaben machen, die auf Ihre Person schließen lassen. Aus diesem Grund wurde der Datensatz nach der Datenerhebung auf faktische Anonymität hin geprüft und es wurden weitere, vorbeugende Maßnahmen gegen De-Anonymisierung getroffen.

- *Entfernung der offenen Angaben:* Laut Ebel und Meyermann (2015, S. 5) „besteht die Gefahr, dass Studienteilnehmer/-innen bei eigentlich unbedenklichen Fragen mit offener Antwortmöglichkeit kritische Informationen preisgegeben haben, die zu einer Identifikation führen könnten“. Aus diesem Grund wurden alle offenen Angaben aus dem Datensatz entfernt und es sind lediglich die (re-)codierten Angaben enthalten.
- *Aggregation von Quasi-Identifikatoren:* Informationen, die in Kombination mit anderen Merkmalen oder durch Anspielung externer Daten geeignet wären eine Person indirekt zu identifizieren, wurden entweder gänzlich entfernt oder mit anderen Informationen zusammengefasst (aggregiert).

Eine detaillierte Übersicht der Merkmale, bei denen Maßnahmen gegen eine De-Anonymisierung getroffen wurden, finden Sie in Tabelle 2.

Tabelle 2: Übersicht der Maßnahmen gegen De-Anonymisierung

Merkmal	SUF
Sonstige Quellen von Forschungsdaten (offen)	Angaben entfernt. Lediglich recodierte Angaben enthalten.
Dauer der Aufbewahrung von Forschungsdaten (offen)	Angaben entfernt. Lediglich in numerische Werte überführte Angaben enthalten.
Weitere Hinderungsgründe bei der Veröffentlichung von Forschungsdaten (offen)	Angaben entfernt. Lediglich codierte Angaben enthalten.
Andere Richtlinien und Standards (offen)	Angaben entfernt.
Sonstige externe Partner (offen)	Angaben entfernt. Lediglich recodierte Angaben enthalten.
Bedarfe zur Verbesserung des Umgangs mit Forschungsdaten (offen)	Angaben entfernt. Lediglich codierte Angaben enthalten.
Andere Art von Software (offen)	Angaben entfernt.
Hochschule (Dropdown-Auswahl)	Angaben entfernt.
Statusgruppe (Einfachauswahl)	Aggregation der Statusgruppen Laboringenieure, Lehrbeauftragte und Hilfskräfte in der Kategorie „anderer Status“

Geschlecht (Einfachauswahl)	Aggregation des diversen Geschlechts in der Kategorie „keine Angabe (Antwortkategorie)“
Anmerkungen (offen)	Angaben entfernt. Lediglich codierte Angaben enthalten.

Die Vorschriften der EU-Datenschutz-Grundverordnung (EU-DSGVO) sowie alle anderen in Deutschland gültigen datenschutzrechtlichen Bestimmungen wurden zu jedem Zeitpunkt der Erhebung eingehalten. Im Rahmen der Verfahrensdokumentation der Kompetenzstelle IT-Doku der Frankfurt UAS wurde das Vorhaben vorab unter anderem vom Datenschutzbeauftragten der Frankfurt UAS geprüft und freigegeben. Zudem fand eine Prüfung des Datensatzes und der Befragungsunterlagen (inkl. der Hinweise zum Datenschutz) durch das FDZ-DZHW statt.

7. Literaturverzeichnis

Bad Wiesseer Kreis (2019). *Lübecker Manifest: 50 Jahre Hochschulen für Angewandte Wissenschaften*. Lübeck. Verfügbar unter: https://www.badwiesseerkreis.de/wp-content/uploads/2019/06/20190614_L%C3%BCbeckerManifest.pdf, letzter Zugriff: 17.07.2023

Ebel, T. & Meyermann, A. (2015). Hinweise zur Anonymisierung von quantitativen Daten. *Forschungsdaten Bildung informiert*. Bd.3. DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation. Frankfurt am Main. <https://doi.org/10.25656/01:21970>

Statistisches Bundesamt (2021). *Personal an Hochschulen*. Fachserie 11. Reihe 4.4 – 2020

Statistisches Bundesamt (2023). *Systematik der Fächergruppen, Lehr- und Forschungsbereiche und Fachgebiete*. Berichtsjahr 2021

Verbund Forschungsdaten Bildung (2019): Hinweise zur Codierung fehlender Werte in der Aufbereitung quantitativer Daten. Version 1.0, *fdbinfo* Nr. 6. <https://doi.org/10.25656/01:22303>

Werth, R. (2023). Generierte Variablen im EVER_FDM Datensatz. *FZDW Technical Reports, 2*. Forschungszentrum Demografischer Wandel (FZDW), Frankfurt University of Applied Sciences. Frankfurt am Main