

September 2025

Judith Hartstein | Clemens Blümel | Daniel Klein

# Trendumfrage Forschungsdaten- infrastrukturen 2024

Daten- und Methodenbericht

Dieses Werk steht unter der Creative Commons Namensnennung – Nicht kommerziell – Weitergabe unter gleichen Bedingungen 3.0 Deutschland Lizenz (CC-BY-NC-SA)

<https://creativecommons.org/licenses/by-nc-sa/3.0/de/>



#### **Bibliographische Angaben**

Hartstein, J., Blümel, C. & Klein, D. (2025). Trendumfrage Forschungsdateninfrastrukturen 2024. Daten- und Methodenbericht. Hannover: Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW).

<https://doi.org/10.21249/dzhw:base4nfdi-dmr:2.0.0>

#### **Projektleitung**

Clemens Blümel

Telefon +49 30 2064177-31

E-Mail: [bluemel@dzhw.eu](mailto:bluemel@dzhw.eu)

#### **Projektmitarbeiter\*innen**

Judith Hartstein

#### **Impressum**

##### **Herausgegeben von**

Deutsches Zentrum für Hochschul- und  
Wissenschaftsforschung GmbH (DZHW)

Lange Laube 12 | 30159 Hannover | [www.dzhw.eu](http://www.dzhw.eu)

Postfach 2920 | 30029 Hannover

Tel.: +49 511 450670-0 | Fax: +49 511 450670-960

##### **Geschäftsführung**

Dr. Marcus Beiner

Steffen Bruckner

##### **Vorsitzender des Aufsichtsrats**

Ministerialdirigent Peter Greisler

##### **Registergericht**

Amtsgericht Hannover | B 210251

Umsatzsteuer-Identifikationsnummer:

DE291239300

September 2025

# Inhaltsverzeichnis

Tabellen-/Abbildungsverzeichnis	I
<b>1 Studienüberblick</b>	<b>3</b>
<b>2 Methoden</b>	<b>4</b>
2.1 Stichprobenziehung .....	4
2.2 Feldphase und Rücklauf .....	4
2.3 Samplebeschreibung .....	4
<b>3 Erhebungsinstrument</b>	<b>7</b>
3.1 Themenübersicht .....	7
3.2 Item-Typen .....	8
<b>4 Datenaufbereitung</b>	<b>11</b>
4.1 Löschung von Fällen .....	11
4.2 Datenstruktur und Datenformat .....	11
4.3 Codierung fehlender Werte .....	13
4.4 Codierung, Recodierung und Erstellung von Variablen .....	14
4.5 Erstellung „Softwaredatensatz“ .....	15
<b>5 Anonymisierung</b>	<b>18</b>
<b>6 Literatur</b>	<b>20</b>
<b>7 Anhang</b>	<b>21</b>
7.1 Einladungsschreiben .....	21
7.2 Informationsblatt zum Datenschutz .....	22
7.3 Anonymisierte Variablen .....	24
7.4 Dokumentation der Herkunft sekundär genutzter Fragen .....	25

## Tabellen-/Abbildungsverzeichnis

Abbildung 1: Einrichtungszuordnung der Befragten der Umfrage FDI 2024 im Vergleich zur Wissenschaftsbefragung 2023 .....	5
Abbildung 2: Statusgruppe der Befragten der Umfrage FDI 2024 im Vergleich zur Wissenschaftsbefragung 2023 .....	5
Abbildung 3: Fachgebiete und Wissenschaftsbereiche der Befragten der Umfrage FDI 2024 im Vergleich zur Wissenschaftsbefragung 2023 .....	6

Abbildung 4: Geschlecht der Befragten der Umfrage FDI 2024 im Vergleich zur Wissenschaftsbefragung 2023 ..... 6

Tabelle 1: Themen der Umfrage FDI 2024 mit Angabe der zugehörigen Items ..... 7

Abbildung 5: Darstellung eines Items vom Typ „Beschreibung“ in der Online-Umfrage ..... 8

Abbildung 6: Darstellung eines Items vom Typ „Checkbox“ in der Online-Umfrage..... 8

Abbildung 7: Darstellung eines Items vom Typ „Radiobutton“ in der Online-Umfrage ..... 9

Abbildung 8: Darstellung eines Items vom Typ „Matrixfrage“ in der Online-Umfrage ..... 9

Abbildung 9: Darstellung eines Items vom Typ „Slider-Element“ in der Online-Umfrage..... 9

Abbildung 10: Darstellung eines Items vom Typ „Freitextfeld“ in der Online-Umfrage.....10

Abbildung 11: Darstellung eines Items vom Typ „Kombinationsfrage“ in der Online-Umfrage...10

Tabelle 2: Codierung fehlender Werte in der Umfrage FDI 2024..... 13

Tabelle 3: Software und digitale Plattformen: Nutzung, Programmieren und Veröffentlichen (Anzahl) 16

Tabelle 4: Softwarenamen und Lizenzangaben vor und nach Anonymisierung (Anzahl) .....17

Tabelle 5: Zugangswege im FDZ-DZHW..... 18

Tabelle 6: Anonymisierte Variablen je Zugangsweg..... 24

Tabelle 7: Dokumentation der Herkunft sekundär genutzter Fragen ..... 25

# 1 Studienüberblick

Die Trendumfrage Forschungsdateninfrastrukturen 2024 (Umfrage FDI 2024) ist Teil der Begleitforschung zu den Nationalen Forschungsdateninfrastruktur (NFDI)-Basisdiensten (Base4NFDI). Die Umfrage FDI 2024 erfasst die Wahrnehmung, Nutzung und Bewertung bestehender und neuer Dateninfrastrukturen und Dienste in der deutschen Forschungslandschaft. Der Fokus liegt auf der Perspektive der (potenziellen) Nutzenden.

Das Projekt wird als Teil der NFDI gefördert von der Deutschen Forschungsgemeinschaft (DFG). Grant Number 521463400.

## 2 Methoden

### 2.1 Stichprobenziehung

Die Umfrage FDI 2024 richtete sich an Wissenschaftler\*innen an deutschen Hochschulen und Forschungseinrichtungen, die (potenziell) Forschungsdateninfrastrukturen nutzen. Die Stichprobenziehung erfolgte zweistufig: Zunächst wurden Kontaktdaten von Wissenschaftler\*innen aus öffentlich zugänglichen Webseiten deutscher Universitäten und gleichgestellter Hochschulen mit Promotionsrecht recherchiert und daraus eine disproportional geschichtete Zufallsstichprobe von 105.655 Adressen gezogen (ausführlich: Fabian et al. 2025a). Im zweiten Schritt erfolgte eine einfache Zufallsauswahl eines Drittels dieser Adressen.

### 2.2 Feldphase und Rücklauf

Zwischen dem 2. und 4. Mai 2024 wurden 35.567 Einladungsschreiben mit personalisiertem Link zum Online-Fragebogen per E-Mail versandt (Anhang 7.1). Es wurden keine Erinnerungsschreiben verschickt. Die Umfrage wurde nach etwas mehr als 10 Wochen am 16. Juni 2024 geschlossen.

Im Umfragezeitraum haben 1.033 Personen die Umfrage begonnen, die Datenschutzerklärung akzeptiert und mindestens eine inhaltliche Frage beantwortet. Insgesamt haben 833 Personen alle Fragen angesehen und die Umfrage bis zum Abschlussbildschirm durchlaufen. Definiert man als Minimalkriterium für teilausgefüllte Fragebögen („partial“) die Beantwortung einer Frage, liegt die (unbereinigte) Rücklaufquote (RR1; AAPOR 2023) bei 3 Prozent

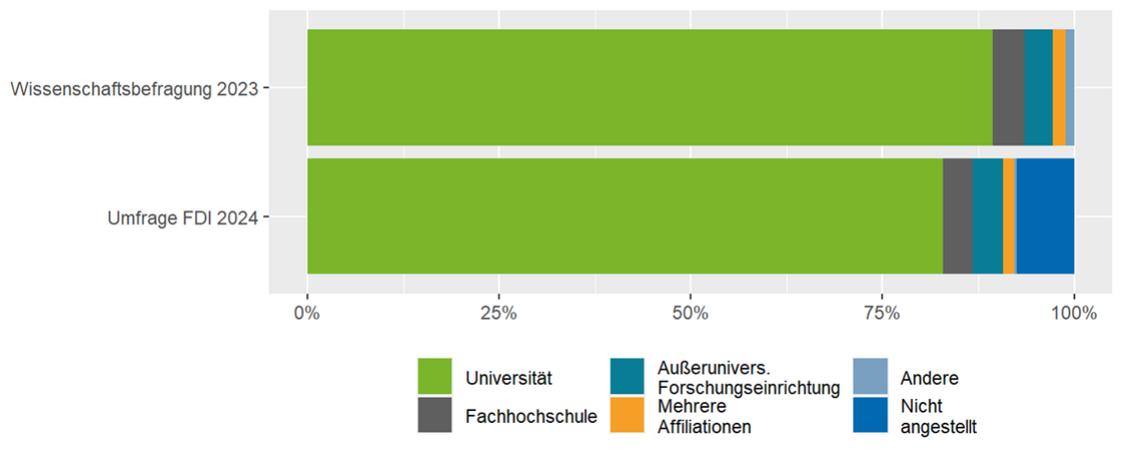
### 2.3 Samplebeschreibung

Zur Beschreibung der Befragten in der Stichprobe bietet sich der Vergleich mit der DZHW-Wissenschaftsbefragung 2023 (Fabian et al. 2025b) an, die auf dem gleichen Auswahlrahmen beruht. Soziodemographische und wissenschaftsbiographische Angaben in der Umfrage FDI 2024 basieren zudem auf den gleichen Items wie bei der Wissenschaftsbefragung 2023 (Anhang 7.4, Tabelle 7). Die Wissenschaftsbefragung 2023 strebt eine repräsentative Stichprobenziehung hauptberuflich wissenschaftlicher Beschäftigter an deutschen Universitäten und gleichgestellten Hochschulen an und erreichte 11.371 Befragte.

#### **Einrichtungszuordnung**

Die Verteilung der Befragten auf unterschiedliche Einrichtungstypen (Universität, Fachhochschule, außeruniversitäre Forschungseinrichtungen und andere) entsprach bei der Umfrage FDI 2024 weitgehend der erwarteten Verteilung aus der Wissenschaftsbefragung 2023 (siehe Abbildung 1). Allerdings wurden Personen ohne Beschäftigungsverhältnis an einer deutschen Hochschule oder Forschungseinrichtung bei der Wissenschaftsbefragung 2023 durch Filterführung ausgeschlossen. In der Trendumfrage Forschungsdateninfrastruktur 2024 macht diese Gruppe 10 Prozent der Befragten aus.

Abbildung 1: Einrichtungszuordnung der Befragten der Umfrage FDI 2024 im Vergleich zur Wissenschaftsbefragung 2023



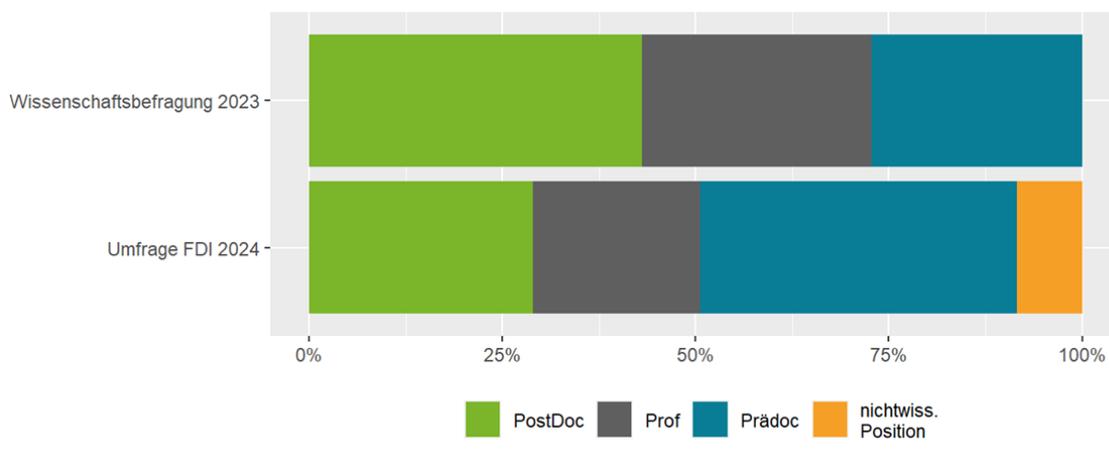
### Statusgruppen

Die Verteilung auf die Statusgruppen (Professor\*innen, Post- und Prädoktorand\*innen) lässt vermuten, dass die Umfrage FDI 2024 im Vergleich zur Wissenschaftsbefragung 2023 mehr Wissenschaftler\*innen vor der Promotion erreicht hat (Abbildung 2). Allerdings wurde die Unterscheidung zwischen Post- und Prädoktorand\*innen erst spät im Fragebogen erhoben. Da einige Teilnehmende die Umfrage vor dieser Frage abgebrochen haben, dürfte die Zahl der Postdoktorand\*innen unterschätzt sein.

Der Anteil der Professor\*innen fällt in der Umfrage FDI 2024 im Vergleich zur Wissenschaftsbefragung 2023 geringer aus. Da die Frage nach einer Professur direkt zu Beginn der Befragung gestellt wurde, ist dieses Ergebnis kein Artefakt systematischer Verzerrungen.

Die Gruppe der Personen ohne wissenschaftliche Position wurde bei der Wissenschaftsbefragung durch Filterführung ausgeschlossen. In der Umfrage FDI 2024 macht diese Gruppe 10 Prozent der Befragten aus.

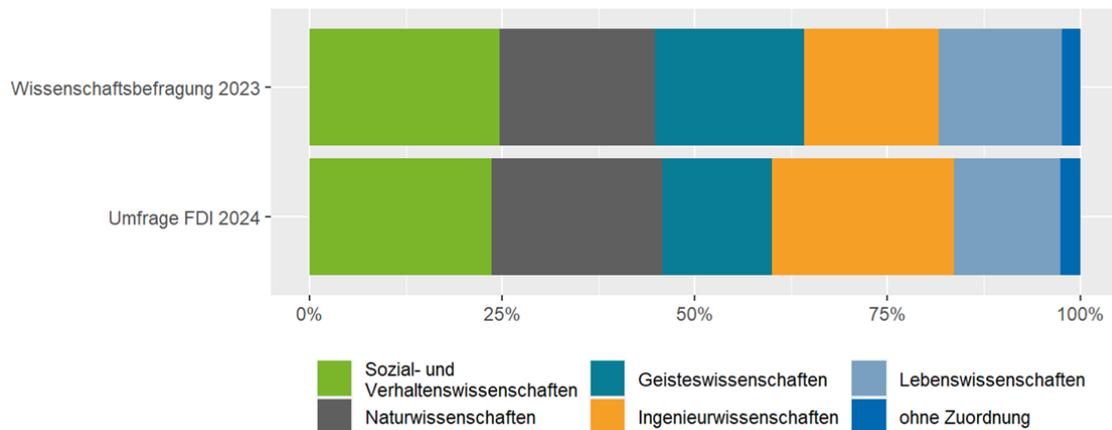
Abbildung 2: Statusgruppe der Befragten der Umfrage FDI 2024 im Vergleich zur Wissenschaftsbefragung 2023



## Fachgebiete und Wissenschaftsbereiche

In der Umfrage FDI 2024 sind die Ingenieur- und Naturwissenschaften stärker vertreten als es angesichts der Wissenschaftsbefragung 2023 zu erwarten wäre. Umgekehrt sind die Geistes-, Lebens- sowie die Sozial- und Verhaltenswissenschaften weniger repräsentiert (Abbildung 3).

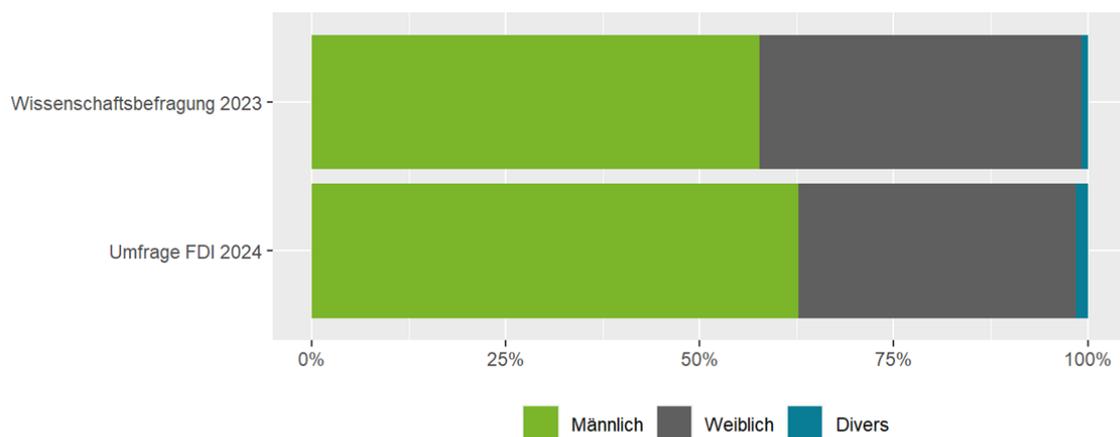
Abbildung 3: Fachgebiete und Wissenschaftsbereiche der Befragten der Umfrage FDI 2024 im Vergleich zur Wissenschaftsbefragung 2023



## Geschlecht

In der Umfrage FDI 2024 zeigt sich im Vergleich zur Wissenschaftsbefragung 2023 eine leichte Überrepräsentation männlicher Befragter. Die (absolute) Zahl der Befragten, die „divers“ als Geschlecht angegeben haben, ist insgesamt gering. Deshalb sollte die Verdoppelung ihres Anteils in der Umfrage FDI 2024 vorsichtig interpretiert werden (Abbildung 4).

Abbildung 4: Geschlecht der Befragten der Umfrage FDI 2024 im Vergleich zur Wissenschaftsbefragung 2023



## 3 Erhebungsinstrument

### 3.1 Themenübersicht

Die Umfrage FDI 2024 umfasste 17 thematische Fragen sowie 10 Fragen zur Wissenschaftsbiografie und Soziodemografie (Tabelle 1).

**Tabelle 1: Themen der Umfrage FDI 2024 mit Angabe der zugehörigen Items**

Thema	Items
Soziodemographie I (5 Items)	Einrichtungstyp Statusgruppe Fachgebiet Epistemische Praktiken I Epistemische Praktiken II
Forschungsdateninfrastrukturen (6 Items)	FDI im Forschungsalltag Diskurs zu FDI Nutzung von FDI Bedarf an FDI Erwartungen an die Weiterentwicklung von FDI Unterstützungsbedarfe zu FDI
NFDI und Base4NFDI (6 Items)	Bekanntheit der NFDI Bekanntheit von Konsortien I Bekanntheit von Konsortien II Dynamik in der Bekanntheit der NFDI Bekanntheit der NFDI-Basisdienste I Bekanntheit der NFDI-Basisdienste II
Software (5 Items)	Nutzung von Software I Auswahl von Software Programmierung von Software Veröffentlichen von Software Nutzung von Software II
Soziodemographie II (5 Items)	Akademischer Lebenslauf Befristung des Arbeitsvertrags Umfang der Lehrverpflichtung Geschlecht Geburtsjahr

Die wissenschaftsbiografischen und soziodemographischen Fragen wurden in zwei Blöcke aufgeteilt, die die inhaltlichen Blöcke einrahmen. Um sicherzustellen, dass zentralen Themen der Wissenschaftsforschung möglichst vollständig erfasst werden, wurden Fragen zur alltäglichen wissenschaftlichen Arbeit an den Anfang gestellt. Für Angaben zum wissenschaftlichen Werdegang, zu arbeitsvertraglichen Bedingungen sowie zu Geschlecht und Alter wurde hingegen ein höherer Anteil fehlender Antworten durch Befragungsabbrüche in Kauf genommen. Die entsprechenden Fragen wurden erst am Ende des Fragebogens gestellt.

Die thematisch inhaltlichen Fragen beziehen sich auf die drei Bereiche

1. Forschungsdateninfrastrukturen
2. NFDI und NFDI-Basisdienste
3. Softwarepraktiken

## 3.2 Item-Typen

Die Daten der Umfrage FDI 2024 wurde mit einem standardisierten Online-Fragebogen in deutscher Sprache erhoben. Die technische Umsetzung erfolgte mit der Umfrage-Software Voxco Online ([www.voxco.com](http://www.voxco.com)). Im Folgenden werden die verwendeten Item-Typen kurz erklärt und beispielhaft dargestellt.

### Item-Typ: Beschreibung

Ein Beschreibungsfeld enthält Informationen für die Teilnehmenden. Ein aktives Element ist nicht enthalten.

Abbildung 5: Darstellung eines Items vom Typ „Beschreibung“ in der Online-Umfrage

**Im Folgenden interessieren wir uns für Ihre Perspektive auf Forschungsdateninfrastrukturen. Damit meinen wir Einrichtungen, welche die Weitergabe von Forschungsdaten an Wissenschaftler\*innen organisieren. Beispiele für solche Infrastrukturen sind Online-Repositorien, digitalisierte Sammlungen oder Forschungsdatenzentren.**

### Item-Typ: Checkbox

In einem Checkbox-Item werden verschiedene Antworten zu einer Frage angeboten. Die Mehrfachauswahl ist grundsätzlich möglich, kann aber eingeschränkt werden. In der Umfrage FDI 2024 konnten Antwortverweigerung (im Beispiel: „Keine Angabe“) nicht mit anderen Auswahlmöglichkeiten kombiniert werden. Teilweise wurden Freitextfelder zu einzelnen Antworten ergänzt (nicht im Beispiel).

Abbildung 6: Darstellung eines Items vom Typ „Checkbox“ in der Online-Umfrage

**Welche Rolle spielen Forschungsdateninfrastrukturen in Ihrer Arbeit?**

- Für meine Arbeit nutze ich Forschungsdateninfrastrukturen.
- Ich stelle Daten mittels Forschungsdateninfrastrukturen zur Verfügung.
- Ich entwickle oder betreue Forschungsdateninfrastrukturen mit.
- Forschungsdateninfrastrukturen spielen für meine Arbeit keine Rolle.
- Keine Angabe

### Item-Typ: Radiobutton

In einem Item vom Typ Radiobutton werden verschiedene Antworten zu einer Frage angeboten. Es kann nur eine Antwort ausgewählt werden. In der Umfrage FDI 2024 wurden teilweise Freitextfelder zu einzelnen Antworten ergänzt (nicht im Beispiel).

Abbildung 7: Darstellung eines Items vom Typ „Radiobutton“ in der Online-Umfrage

**Kennen Sie die Nationale Forschungsdateninfrastruktur (NFDI)?**

- Die NFDI ist mir unbekannt.
- Ich habe von der NFDI bereits gehört.
- Ich kenne mich mit Zielen und Verfahrensweisen der NFDI gut aus.
- Keine Angabe

### Item-Typ: Matrixfrage

In einer Matrixfrage werden verschiedene Teilfragen zu einem Thema kombiniert. Es kann je Zeile nur eine Antwort ausgewählt werden.

Abbildung 8: Darstellung eines Items vom Typ „Matrixfrage“ in der Online-Umfrage

**Wie schätzen sie die folgenden Aussagen über Sie selbst und ihr Fachgebiet ein?**

	trifft nicht zu		trifft voll zu	
In meinem Fach gibt es einen Konsens über die Archivierung und das Teilen von Daten in Forschungsdateninfrastrukturen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Für mein Fach gibt es umfangreiche Handreichungen und Richtlinien für die Nutzung von Forschungsdateninfrastrukturen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
In meinem Fach werden Forschungsdateninfrastrukturen bisher nicht diskutiert.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich interessiere mich persönlich wenig für Forschungsdateninfrastrukturen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

### Item-Typ: Slider-Element

In einem Slider-Element können Befragte ihre Einstellung zu einem Thema graduell ausdrücken. Der Regler kann dafür nach links und rechts verschoben werden. In der Umfrage FDI 2024 wurde die mittlere Ausprägung als Voreinstellung gewählt.

Abbildung 9: Darstellung eines Items vom Typ „Slider-Element“ in der Online-Umfrage

**Reicht das Angebot an Forschungsdateninfrastrukturen für Ihre Disziplin Ihrer Meinung nach aus?**

Nein, ist viel zu wenig  Ja, reicht voll aus

### Item-Typ: Freitextfeld

In einem Freitextfeld können Befragte textbasierte und numerische Antworten frei eingeben. Die Größe des Feldes in der Darstellung sowie die zulässige Zeichenzahl kann für jede Frage separat eingestellt werden. In der Umfrage FDI 2024 wurden Eingabemöglichkeiten auf einen plausiblen Wertebereich eingeschränkt (z. B. Stunden pro Woche). Freitextfelder wurden teilweise mit einem Radiobutton für die explizite Antwortverweigerung (im Beispiel: „keine Angabe“) ergänzt.

Abbildung 10: Darstellung eines Items vom Typ „Freitextfeld“ in der Online-Umfrage

**Welche Erwartungen haben Sie als Nutzer\*in an die (Weiter-)entwicklung von Forschungsdateninfrastrukturen für Ihren Fachbereich?**

Keine Angabe

### Item-Typ: Kombinationsfeld

In einem Kombinationsfeld sind verschiedene Elemente kombiniert. In der Umfrage FDI 2024 wurde ein solches Feld nur für die Frage nach dem akademischen Lebenslauf genutzt. Dieses Item ist zeilenweise aufgebaut und fragt im konkreten Beispiel nach einer Jahreszahl mittels Freitextfeld (numerisch, vierstellig), sowie nach einem Land aus einer hinterlegten Liste, welches per Drop-Down ausgewählt werden soll.

Abbildung 11: Darstellung eines Items vom Typ „Kombinationsfrage“ in der Online-Umfrage

**Zu weiteren Etappen Ihres akademischen Lebenslaufs:**

**In welchem Jahr und in welchem Land haben Sie die folgenden Etappen absolviert?**

	Jahr	Land
(Erster) Hochschulabschluss (auf Master-/Diplomniveau)	<input type="text"/>	Select an answer... ▾
(Erste) Promotion	<input type="text"/>	Select an answer... ▾
Erstberufung	<input type="text"/>	Select an answer... ▾

# 4 Datenaufbereitung

Zunächst wurde aus Voxco ein Rohdatensatz im XLSX-Format exportiert. Die weitere Aufbereitung und Anonymisierung (Abschnitt 5) zur Erstellung der Scientific Use Files (SUF) erfolgten mit R (R Core Team 2024) und Stata (StataCorp. 2023).

## 4.1 Löschung von Fällen

Im Rohdatensatz wurde für alle 35.567 eingeladenen Befragten eine Zeile angelegt. In den SUF wurden nur die Angaben der 1.033 befragten Personen behalten, die mindestens eine Frage nach der Datenschutzerklärung beantwortet haben (teilausgefüllte Fragebögen). Weitere Ausschlüsse erfolgten nicht.

## 4.2 Datenstruktur und Datenformat

Die Daten der Umfrage FDI 2024 werden als SUF in zwei Datensätzen bereitgestellt:

1. „Personendatensatz“ (1.033 Beobachtungen × 97 Variablen)
2. „Softwaredatensatz“ (4.857 Beobachtungen × 7 Variablen)

Der „Personendatensatz“ liegt im breiten (wide) Datenformat vor, in dem alle Informationen zu einer befragten Person in einer Zeile erfasst werden. Eine Zeile im „Personendatensatz“ entspricht also genau einer befragten Person.

Der „Softwaredatensatz“ liegt im langen (long) Datenformat vor, in dem Informationen zu einer befragten Person über mehrere Zeilen verteilt sind. Eine Zeile im „Softwaredatensatz“ erfasst den Umgang einer Person mit einer spezifischen Software bzw. digitalen Plattform im Forschungsprozess (Abschnitt 4.5). Für 691 befragte Personen liegen jeweils zwischen 1 und 30 Zeilen vor.

Der „Personendatensatz“ und der „Softwaredatensatz“ werden jeweils in drei Dateiformaten bereitgestellt:

1. Stata (.dta)
2. SPSS (.sav)
3. CSV (.csv)

Alle SUF basieren auf dta-Dateien im Stata-Format und werden aus Stata in die jeweiligen Dateiformate SPSS und CSV exportiert.

Die Reihenfolge der Variablen orientiert sich in allen Dateiformaten an der Position der entsprechenden Fragen im Erhebungsinstrument (vgl. Variablenfragebogen). Die Variablennamen sind in allen Dateiformaten identisch. Im CSV-Format stehen die Variablennamen im Datensatz in der ersten Zeile; in den Zuordnungstabellen („dictionaries“) stehen die Variablennamen in der ersten Spalte.

Die Daten werden in allen Dateiformaten in deutscher und englischer Sprache bereitgestellt.

### Stata-Format (.dta)

Nominale, nominal binäre und ordinale Variablen sind im Stata-Format als numerische Variablen mit Wertelabels angelegt.

Metrische Variablen sind im Stata-Format ebenfalls als numerische Variablen angelegt. Für fehlende Werte sind entsprechende Wertelabels hinterlegt.

Deutsche und englische Labels sind in einem Datensatz integriert. In Stata kann mit dem Befehl `label language` zwischen den Sprachen gewechselt werden.

Das Stata Dateiformat umfasst je Zugangsweg („o“, „r“, oder „d“; Abschnitt 5) zwei Dateien:

- `base4nfdi_p_Zugangsweg_2-0-0.dta`
  - „Personendaten“ (Deutsch / Englisch)
- `base4nfdi_e_Zugangsweg_2-0-0.dta`
  - „Software-daten“ (Deutsch / Englisch)

### SPSS-Format (.sav)

Nominale, nominal binäre und ordinale Variablen sind im SPSS-Format als numerische Variablen mit Wertelabels angelegt.

Metrische Variablen sind im SPSS-Format ebenfalls als numerische Variablen angelegt. Für fehlende Werte sind entsprechende Wertelabels hinterlegt.

Für alle Variablen ist standardmäßig das Messniveau „Nominal“ eingestellt.

Deutsche und englische Labels werden in getrennten Datensätzen bereitgestellt.

Das SPSS Dateiformat umfasst je Zugangsweg („o“, „r“, oder „d“; Abschnitt 5) vier Dateien:

- `base4nfdi_p_Zugangsweg_2-0-0_de.sav`
  - „Personendaten“ (Deutsch)
- `base4nfdi_e_Zugangsweg_2-0-0_de.sav`
  - „Software-daten“ (Deutsch)
- `base4nfdi_p_Zugangsweg_2-0-0_en.sav`
  - „Personendaten“ (Englisch)
- `base4nfdi_e_Zugangsweg_2-0-0_en.sav`
  - „Software-daten“ (Englisch)

### CSV-Format (.csv)

Nominale Variablen sind im CSV-Format als Character-Variablen (auch: Zeichenketten, Strings) angelegt. Dies dient zur Vereinfachung von Auswertung und Plausibilitätsprüfung bei der Sekundärdatennutzung, insbesondere mit R.

Ordinale und (nominal) binäre Variablen sind im CSV-Format als Integer-Variablen (Ganzzahlen) angelegt. Um die Visualisierung mit verbalisierten Antworten zu erleichtern, stehen Zuordnungstabellen („dictionaries“) zur Verfügung.

Metrische Variablen sind im CSV-Format als Gleitkommazahl („Float“) angelegt. In den SUF sind allerdings nur Ganzzahlen enthalten. Für metrische Variablen steht naturgemäß keine Verbalisierung zur Verfügung. Für fehlende Werte sind aber verbale Beschreibungen in den Zuordnungstabellen hinterlegt.

Deutsche und englische Werte (Character-Variablen) und Labels werden in getrennten Datensätzen und Zuordnungstabellen („dictionaries“) bereitgestellt.

Das CSV-Dateiformat umfasst je Zugangsweg („o“, „r“, oder „d“; Abschnitt 5) zehn Dateien:

- base4nfdi\_p\_Zugangsweg\_2-0-0\_de.csv
  - „Personendaten“ (Deutsch)
- base4nfdi\_e\_Zugangsweg\_2-0-0\_de.csv
  - „Softwaredaten“ (Deutsch)
- base4nfdi\_variable-labels\_p\_Zugangsweg\_2-0-0\_de.csv
  - Zuordnungstabelle Variablenlabels „Personendaten“ (Deutsch)
- base4nfdi\_value-labels\_p\_Zugangsweg\_2-0-0\_de.csv
  - Zuordnungstabelle Wertelabels „Personendaten“ (Deutsch)
- base4nfdi\_variable-labels\_e\_Zugangsweg\_2-0-0\_de.csv
  - Zuordnungstabelle Variablenlabels „Softwaredaten“ (Deutsch)
- base4nfdi\_p\_Zugangsweg\_2-0-0\_en.csv
  - „Personendaten“ (Englisch)
- base4nfdi\_e\_Zugangsweg\_2-0-0\_en.csv
  - „Softwaredaten“ (Englisch)
- base4nfdi\_variable-labels\_p\_Zugangsweg\_2-0-0\_en.csv
  - Zuordnungstabelle Variablenlabels „Personendaten“ (Englisch)
- base4nfdi\_value-labels\_p\_Zugangsweg\_2-0-0\_en.csv
  - Zuordnungstabelle Wertelabels „Personendaten“ (Englisch)
- base4nfdi\_variable-labels\_e\_Zugangsweg\_2-0-0\_en.csv
  - Zuordnungstabelle Variablenlabels „Softwaredaten“ (Englisch)

### 4.3 Codierung fehlender Werte

Fehlende Werte wurden von der Befragungssoftware Voxco undifferenziert als „NA“ (not available) erfasst. Bei der Datenaufbereitung wurden in den SUF dreistellige negative Werte vergeben. Die verwendete Codierung ist in Tabelle 2 dargestellt.

Tabelle 2: Codierung fehlender Werte in der Umfrage FDI 2024

Code	Wertelabel (Deutsch)	Wertelabel (Englisch)
-999	weiß nicht	don't know
-997	keine Angabe (Antwortkategorie)	no answer (response category)
-969	unbekannter fehlender Wert	unknown missing
-967	anonymisiert	anonymized
-964	unbekannt / nicht anwendbar	unknown / not applicable
-946	ohne Zuordnung	not assigned

Der generische fehlende Wert -969 beinhaltet sowohl übersprungene als auch wegen Abbruch der Befragung nicht erreichte Fragen. Vereinzelt wurden als Anonymisierungsmaßnahme auch gültige Antwortoptionen mit dem generischen fehlenden Wert -969 aggregiert. Der fehlende Wert -967 wurde im Rahmen der Anonymisierung (Abschnitt 5) vergeben.

## 4.4 Codierung, Recodierung und Erstellung von Variablen

### **Eindeutiger Identifikator (Variable: „pid“)**

Zur eindeutigen Identifizierung der Beobachtungen im „Personendatensatz“ steht die Variable „pid“ zur Verfügung. Sie enthält fortlaufende Nummern, die den Beobachtungen nach zufälliger Sortierung zugewiesen wurden. Der „Personendatensatz“ lässt sich über die Variable „pid“ mit dem „Software-datensatz“ verknüpfen.

### **Einrichtungstyp (Variable: „einrichtung“)**

Der Einrichtungstyp, mit dem die Befragten affiliert sind, wurde mittels Checkbox erhoben, sodass Mehrfachantworten möglich waren. Diese Antworten waren im Datensatz ursprünglich in binären Variablen enthalten. Bei der Datenaufbereitung wurde daraus die kategoriale Variable „einrichtung“ erstellt. Dabei wurden Mehrfachantworten als „Mehrere Affiliationen“ codiert, was gleichzeitig der Anonymisierung dient (Abschnitt 5). Die Frage wurde aus der DZHW-Wissenschaftsbefragung 2023 übernommen.

### **Statusgruppe / Beschäftigungsposition (Variable: „posi“)**

Die Statusgruppe der Befragten wurde mittels Radiobutton erhoben. Dabei wurde innerhalb der Gruppe der wissenschaftlichen Mitarbeitenden zunächst keine Differenzierung zwischen Doktorand\*innen und Post-Doktorand\*innen vorgenommen. Es wurde aber ein zusätzliches Freitextfeld angeboten. Bei der Datenaufbereitung wurde zunächst die Beschäftigungsposition anhand der Angaben im Freitextfeld nachcodiert. Anschließend erfolgte eine Aggregation in drei Gruppen: Professor\*in und Juniorprofessor\*in, wissenschaftliche Mitarbeitende, und nichtwissenschaftliche Position. Die Aggregation dient gleichzeitig der Anonymisierung (Abschnitt 5). Zuletzt wurden die wissenschaftlichen Mitarbeiter\*innen anhand der Angabe zum Jahr der Promotion (erhoben im akademischen Lebenslauf, Soziodemographie II) nachträglich der Gruppen Prä- und Postdocs zugeordnet. Im „Personendatensatz“ ist die finale Variable „posi“ verfügbar. Die Frage wurde aus der DZHW-Wissenschaftsbefragung 2023 übernommen.

### **Fachgebiet (Variablen: „dfg10“, „dfg5“)**

Das Fachgebiet, dem sich die Befragten zuordnen, wurde in 49 Kategorien gemäß DFG-Fachsystematik der Wissenschaftsbereiche, Fachgebiete, Fachkollegien und Fächer (Amtsperiode 2020–2024) erhoben.<sup>1</sup> Alternativ konnten die Befragten ihre Antwort in einem Freitextfeld angeben. Bei der Datenaufbereitung wurde zunächst die Fachzuordnung anhand der Angaben im Freitextfeld nachcodiert. Anschließend wurden die Fachgebiete zu 10 bzw. 5 Kategorien zusammengefasst, wobei diese Aggregation zugleich der Anonymisierung dient (Abschnitt 5). Im „Personendatensatz“ sind die entsprechenden Variablen „dfg10“ und „dfg5“ verfügbar. Im „Softwaredatensatz“ ist nur die Variable „dfg5“ enthalten. Die Frage wurde aus der DZHW-Wissenschaftsbefragung 2023 übernommen.

### **Bekanntheit von Konsortien I (Variable: „bekannth\_konsortien“)**

Die Bekanntheit der Konsortien der NFDI wurde mittels Checkbox erhoben, sodass Mehrfachantworten möglich waren. Diese Antworten waren im Datensatz ursprünglich in binären Variablen enthalten. Bei der Datenaufbereitung wurde daraus die kategoriale Variable „bekannth\_konsortien“ erstellt. Um ein ordinales Skalenniveau zu erreichen, wurde bei Mehrfachantworten der höchstgenannte

---

<sup>1</sup> DFG-Fachsystematik der Wissenschaftsbereiche, Fachgebiete, Fachkollegien und Fächer für die Amtsperiode 2020–2024. <https://www.dfg.de/resource/blob/175334/89ba4a3464c99aaea40fdef47367e7b2/fachsystematik-2020-2024-de-grafik-data.pdf> (abgerufen am 25.03.2025)

Bekanntheitsgrad übernommen. Die Aktivität in einem Konsortium wurde dabei nicht berücksichtigt und, sofern dies die einzige Antwort war, zur Anonymisierung (Abschnitt 5) mit dem fehlenden Wert -969 („unbekannter fehlender Wert“) aggregiert.

#### **Bekanntheit der NFDI Basisdienste I (Variable: „bekannth\_base“)**

Die Bekanntheit der Basisdienste der NFDI wurde mittels Checkbox erhoben, sodass Mehrfachantworten möglich waren. Diese Antworten waren im Datensatz ursprünglich in binären Variablen enthalten. Bei der Datenaufbereitung wurde daraus die kategoriale Variable „bekannth\_base“ erstellt. Um ein ordinales Skalenniveau zu erreichen, wurde bei Mehrfachantworten der höchstgenannte Bekanntheitsgrad codiert. Die Beteiligung an der Entwicklung entsprechender Dienste wurde dabei nicht berücksichtigt und, sofern dies die einzige Antwort war, zur Anonymisierung (Abschnitt 5) mit dem fehlenden Wert -969 („unbekannter fehlender Wert“) aggregiert.

#### **Lehrverpflichtung (Variable: „wibef\_fb9\_lehrverpfl“)**

Der Umfang der Lehrverpflichtung wurde über ein numerisches Freitextfeld erhoben. Zusätzlich war es möglich, den Radiobutton „Ich habe keine Lehrverpflichtung“ anzukreuzen. Bei der Datenaufbereitung wurde die Antwort „Ich habe keine Lehrverpflichtung“ mit dem Wert 0 codiert. Im „Personendatensatz“ ist die entsprechende numerische Variable „wibef\_fb9\_lehrverpfl“ verfügbar. Die Frage wurde aus der DZHW-Wissenschaftsbefragung 2023 übernommen.

## **4.5 Erstellung „Softwaredatensatz“**

Die Rolle von Software und digitalen Plattformen im Forschungsprozess wurde in mehreren Fragen erhoben, wobei eine Kombination aus Checkbox bzw. Radiobutton und Freitextfeldern eingesetzt wurde. Neben den geschlossenen Antwortoptionen konnten die Befragten auch Namen spezifischer Software oder digitaler Plattformen angeben, die sie nutzen, programmieren, oder veröffentlichen. Dieses Vorgehen orientiert sich am Fischer-Namensgenerator. Die erhobenen Daten eignen sich zur Netzwerkanalyse (Jansen 1999, S. 76 ff.).<sup>2</sup>

Der „Softwaredatensatz“ enthält die aufbereiteten Angaben zu folgenden Fragen:

- 5.2 „Nutzen Sie Softwareprogramme und Plattformen in Ihrem Forschungs- und Publikationsprozess? Wenn ja, welche?“
- 5.4 „Programmieren Sie als Teil Ihres Forschungsprozesses oder Ihrer Arbeit? Wenn ja, welche Programmiersprachen und Technologien nutzen Sie hierfür überwiegend?“
- 5.5 „Veröffentlichen Sie (auch gelegentlich) Software oder Programmcode? Wenn ja, wo?“

Eine Übersicht der entsprechenden Angaben ist in Tabelle 3 dargestellt. Zur Erstellung des „Softwaredatensatzes“ wurden die Angaben zunächst zusammengefasst und anschließend durch weitere Informationen ergänzt. Im Folgenden wird die Erstellung der entsprechenden Variablen beschrieben.

#### **Software-Engagement (Variable: „category“)**

Die unterschiedlichen Dimensionen der drei Fragen 5.2, 5.4 und 5.5 sind durch die Ausprägungen „Nutzung“, „Programmieren“ und „Veröffentlichen“ von Software bzw. digitalen Plattformen in der Variable „category“ repräsentiert.

---

<sup>2</sup> Aus dieser Perspektive stellt der „Softwaredatensatz“ Beziehungen der Art „Befragte Person X nennt Software Y für Zweck Z“ ein dünn besetztes Netzwerk („sparse network“) dar. Der Datensatz liegt im langen (long) Format als Kantenliste („edge list“) vor.

### Art der Software / Ort der Veröffentlichung (Variable: „type“)

Die geschlossenen Antwortoptionen der Fragen 5.2, 5.4 und 5.5 sind in der kategorialen Variablen „type“ zusammengefasst.

### Name der Software / Plattform (Variable: „name\_anon“)

Die offenen Angaben der Namen spezifischer Software bzw. Plattformen wurden in mehreren Schritten codiert und teilweise anonymisiert.

1. **Automatische Zerlegung:** Mehrfachnennungen von Softwarenamen in den offenen Angaben wurden zunächst automatisiert an den Kommas getrennt.
2. **Manuelle Korrektur der Zerlegung:** Die automatisch identifizierten Softwarenamen wurden anschließend mit den unbearbeiteten offenen Angaben abgeglichen und bei Bedarf manuell korrigiert.
3. **Vereinheitlichung der Schreibweise und Zusammenfassung:** Unterschiedliche Schreibweisen der gleichen Software wurden manuell vereinheitlicht. Dabei wurden insbesondere Versionsbezeichnungen entfernt. Inhaltlich ähnliche, aber unspezifische Nennungen wie „Bibliothekskataloge“ oder „selbstentwickelt“ wurden zu gemeinsamen Kategorien zusammengefasst.
4. **Anonymisierung:** Um eine mögliche Re-Identifizierung von Befragten anhand selten genannter oder selbstentwickelter Software auszuschließen, wurden Softwarenamen, die seltener als vier Mal genannt wurden, in der Kategorie „andere“ Software zusammengefasst.

Die vereinheitlichten und anonymisierten Namen der genannten Software bzw. Plattformen sind in der Variable „name\_anon“ enthalten.

Tabelle 3: Software und digitale Plattformen: Nutzung, Programmieren und Veröffentlichen (Anzahl)

Fragenummer und Antwortoption	Genannt	Offene Angabe	Beobachtungen im SUF
5.2			
Textverarbeitung/Literaturverwaltung	782	536	1105
Datenerhebung/Analyse/Simulation	590	409	1016
Wissensorganisation/Suchmaschinen	577	340	700
Kommunikation/Projektmanagement	325	213	395
Sonstige Software (Nutzung)	90	65	115
5.4			
Programmiersprachen	472	435	984
weitere Technologien	100	85	172
5.5			
an folgenden Orten (Veröffentlichen Software)	274	263	370
Spaltensumme	3210	2346	4857

### Art der Softwarelizenz (Variable: „license\_category“)

Die in den offenen Angaben genannte Software bzw. Plattformen wurden in zwei Schritten durch Informationen zur Offenheit der entsprechenden Lizenz ergänzt.

1. **Automatisierte Anreicherung mit Informationen zur Offenheit:** Zunächst wurde Lizenzinformationen über die Wikidata-API ([www.wikidata.org](http://www.wikidata.org)) abgefragt. Allerdings konnten mehr als die Hälfte der genannten Softwarenamen keine Lizenzinformationen zugeordnet werden (Tabelle 4). Eine eindeutige Zuordnung war zudem häufig nicht möglich, etwa weil die Lizenzoffenheit von der jeweiligen Softwareversion abhängt.
2. **Manuelle Ergänzung von Informationen zur Offenheit:** Aufgrund der Limitationen der automatisierten Anreicherung wurden Informationen zur Lizenzoffenheit durch manuelle Webrecherchen ergänzt.

Die Variable „license\_category“ unterscheidet zwischen „frei[en]/offen[en]“ und „geschlossen[en]/proprietär[en]“ Softwarelizenzen. Für Software, zu der keine eindeutigen Lizenzinformationen ermittelt werden konnten, wurde die Lizenz als „unbekannt/nicht anwendbar“ codiert.

Tabelle 4: Softwarenamen und Lizenzangaben vor und nach Anonymisierung (Anzahl)

Rohdaten	davon in Wikidata	davon eindeutige Lizenzinformationen	SUF („Softwaredatensatz“)
875	331	330	190 (inkl. „andere“)

### AusgangsvARIABLE (Variable: „questionnaire\_item“)

Die Variable „questionnaire\_item“ enthält für jede Zeile des „Softwaredatensatzes“ den Variablennamen, aus der die jeweiligen Informationen stammen.

## 5 Anonymisierung

**[Datenschutzrechtlicher Rahmen]** Für personenbezogene Daten<sup>3</sup>, die in freiwilligen Befragungen durch das DZHW erhoben werden, gelten die EU-Datenschutz-Grundverordnung (EU-DSGVO) und das Bundesdatenschutzgesetz in seiner Neufassung vom 30. Juni 2017.<sup>4</sup> Danach sind personenbezogene Daten für die Weitergabe zur wissenschaftlichen Sekundärnutzung (ohne Vorliegen einer Einverständniserklärung zur Sekundärnutzung der personenbezogenen Daten) in der Regel derart aufzubereiten, dass „die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können, sofern diese zusätzlichen Informationen gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, die gewährleisten, dass die personenbezogenen Daten nicht einer identifizierten oder identifizierbaren natürlichen Person zugewiesen werden“ können (Art. 4 Abs. 5 DSGVO; s. auch Art. 89 DSGVO sowie Erwägungsgrund 26 DSGVO).

**[Sicherstellung des Datenschutzes im FDZ-DZHW]** Im FDZ-DZHW wird der Datenschutz der Befragten über rechtlich-organisatorische Maßnahmen, technische Maßnahmen und statistische Maßnahmen sichergestellt<sup>5</sup>, die in unterschiedlicher Weise kombiniert werden können. Aus den Kombinationen ergeben sich vier potenzielle Zugangswege, die in Tabelle 5 im Überblick dargestellt sind.<sup>6</sup>

Tabelle 5: Zugangswege im FDZ-DZHW

	Campus Use File (CUF): Download	Scientific Use File (SUF): Download	Scientific Use File (SUF): Remote-Desktop	Scientific Use File (SUF): On-Site
rechtlich-organisatorische Maßnahmen	sehr gering	gering	moderat	hoch
technische Maßnahmen	sehr gering	gering	moderat	hoch
statistische Maßnahmen	sehr hoch	hoch	moderat	gering

Je stärker der Datenzugang rechtlich-organisatorisch und technisch kontrolliert und beschränkt wird, desto geringer ist das Risiko einer De-Anonymisierung der Daten. In der Folge müssen die Daten umso weniger mittels statistischer Maßnahmen anonymisiert, also um Informationen reduziert, werden – das heißt, desto größer bleibt ihr Analysepotenzial.

<sup>3</sup> „Personenbezogene Daten“ (sind) alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person (im Folgenden ‚betroffene Person‘) beziehen; als identifizierbar wird eine natürliche Person angesehen, die direkt oder indirekt, insbesondere mittels Zuordnung zu einer Kennung wie einem Namen, zu einer Kennnummer, zu Standortdaten, zu einer Online-Kennung oder zu einem oder mehreren besonderen Merkmalen identifiziert werden kann, die Ausdruck der physischen, physiologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität dieser natürlichen Person sind“ (Art. 4 Abs. 1 DSGVO).

<sup>4</sup> Die DSGVO gilt grundsätzlich innerhalb der EU und somit ebenfalls für das DZHW. Das BDSG in seiner Neufassung vom 30. Juni 2017 (Gesetz zur Anpassung des Datenschutzrechts an die Verordnung (EU) 2016/679 und zur Umsetzung der Richtlinie (EU) 2016/680 (Datenschutz-Anpassungs- und Umsetzungsgesetz EU DSAnpUG-EU)) kommt teils zusätzlich zur Anwendung, da die DZHW GmbH juristisch als öffentliche Stelle des Bundes betrachtet wird (§ 2 Abs. 3 BDSG). Der Bund hält die absolute Mehrheit der Anteile der DZHW GmbH und das Institut erfüllt Aufgaben der öffentlichen Verwaltung des Bundes im weitesten Sinn.

<sup>5</sup> Das Datenschutzkonzept des FDZ-DZHW ist angelehnt an den Portfolio-Ansatz von Lane, Heus und Mulcahy (2008, 6ff.), an dem sich bereits das Leibniz-Institut für Bildungswissenschaften (LifBi) (vgl. Koberg 2016, 699ff.) und das FDZ der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (vgl. Hochfellner et al. 2012, 9f.) orientieren.

<sup>6</sup> Für genauere Informationen vgl. <https://www.fdz.dzhw.eu/de/datennutzung>

Die Daten der Umfrage FDI 2024 werden über die Zugangswege „Scientific Use File (SUF): Download“, „Scientific Use File (SUF): Remote-Desktop“ und „Scientific Use File (SUF): On-Site“ bereitgestellt. Im Folgenden werden die konkreten durchgeführten statistischen Anonymisierungsmaßnahmen genauer erläutert.

**[Statistische Anonymisierungsmaßnahmen]** Im Rahmen der statistischen Anonymisierungsmaßnahmen wurden zunächst alle Informationen darauf geprüft, ob sich über sie Personen direkt identifizieren lassen. Diese direkten Identifikatoren, wie Namen und E-Mail-Adressen, wurden bereits während der Feldphase in einem separaten Datensatz verarbeitet und sind somit per se nie in den Befragungsdaten enthalten gewesen. Die Original-Identifikationsnummer wurde durch das FDZ entfernt und durch eine neue zufällig vergebene Identifikationsnummer ersetzt.

Darüber hinaus wurde der Empfehlung von Ebel und Meyermann gefolgt, offene Angaben zu löschen „selbst wenn die jeweiligen Fragestellungen an sich unproblematisch sind. Denn es besteht die Gefahr, dass Studienteilnehmer/-innen bei eigentlich unbedenklichen Fragen mit offener Antwortmöglichkeit kritische Informationen preisgegeben haben, die zu einer Identifikation führen könnten“ (Ebel & Meyermann 2015, S. 5).

Anschließend wurden die Quasi-Identifikatoren bestimmt, also Informationen, die in Kombination oder durch die Anspielung externer Informationen geeignet sind, eine Person indirekt zu identifizieren. Um eine eindeutige Zuordnung der Daten zu unterbinden, werden diese Schlüsselmerkmale nur in aggregierter Form freigegeben oder komplett anonymisiert.

Zuletzt wurde geprüft, ob in den Daten sensible Informationen, z. B. zur Gesundheit, sexuellen Orientierung und zu politischen Einstellungen, enthalten waren, für die bei den Befragten kein zusätzliches Einverständnis für die Sekundärnutzung eingeholt wurde. Diese eignen sich zwar nicht notwendig zur Re-Identifikation von Individuen oder Institutionen, jedoch können die Informationen im Falle einer De-Anonymisierung nutzbringend sein (vgl. Koberg 2016, S. 694) und sind daher besonders schützenswert (Art. 9 DSGVO, Erwägungsgrund 51 DSGVO). Die sensiblen Informationen werden – je nach Zugangsweg – nur in aggregierter Form freigegeben oder komplett anonymisiert.

Tabelle 6 (Anhang 7.3) zeigt die im Zuge der Anonymisierung recodierten und neu erstellten Variablen sowie die jeweiligen Zugangswege.

## 6 Literatur

- The American Association for Public Opinion Research (AAPOR) (2023). Standard Definitions. Final Dispositions of Case Codes and Outcome Rates for Surveys. 10th edition. <https://aapor.org/wp-content/uploads/2023/05/Standards-Definitions-10th-edition.pdf> (Zugegriffen am: 20. 07.2023).
- Ebel, T., & Meyermann, A. (2015). Hinweise zur Anonymisierung von quantitativen Daten. Forschungsdaten Bildung informiert. Bd.3. Frankfurt am Main: Deutsches Institut für Internationale Pädagogische Forschung. <https://doi.org/10.25656/01:21970> (Zugegriffen am: 14.10.2024).
- Fabian, G., Heger, C., Just, A. & Weber, A. (2025a). DZHW-Wissenschaftsbefragung 2023. Daten- und Methodenbericht zur DZHW-Wissenschaftsbefragung 2023. Hannover: Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW). <https://doi.org/10.21249/DZHW:scs2023-dmr-de:1.0.1>
- Fabian, G., Heger, C. & Just, A. (2025b). DZHW-Wissenschaftsbefragung 2023. Datenerhebung: 2023. Version: 1.0.1. Datenpaketzugangsweg: Remote-Desktop-SUF. Hannover: FDZ-DZHW. Datenkuratierung: Weber, A. <https://doi.org/10.21249/DZHW:scs2023:1.0.1>.
- Hochfellner, D., Müller, D., Schmucker, A., & Roß, E. (2012). Datenschutz am Forschungsdatenzentrum. (FDZ-Methodenreport 06/2012). Nürnberg: Forschungsdatenzentrum (FDZ) der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (IAB). [https://doku.iab.de/fdz/reporte/2012/MR\\_06-12.pdf](https://doku.iab.de/fdz/reporte/2012/MR_06-12.pdf) (Zugegriffen am: 20.08.2024).
- Jansen, D. (1999). Einführung in die Netzwerkanalyse. Grundlagen, Methoden, Anwendungen. Leske + Budrich.
- Koberg, T. (2016). Disclosing the National Educational Panel Study. In H.-P. Blossfeld, J. v. Maurice, M. Bayer, & J. Skopek (Hrsg.), *Methodological Issues of Longitudinal Surveys. The example of the National Educational Panel Study* (S. 691–708). Wiesbaden: Springer VS. [https://doi.org/10.1007/978-3-658-11994-2\\_38](https://doi.org/10.1007/978-3-658-11994-2_38).
- Lane, J., Heus, P., & Mulcahy, T. (2008). Data access in a cyber world: Making use of cyberinfrastructure. *Transactions on Data Privacy*, 1(1), 2–16.
- R Core Team (2024). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- StataCorp. 2023. Stata Statistical Software: Release 18. College Station, TX: StataCorp LLC.

# 7 Anhang

## 7.1 Einladungsschreiben

Sehr geehrte:r [Anrede],

hiermit laden wir Sie herzlich zur „Trendumfrage Forschungsdateninfrastrukturen 2024“ ein, die am Deutschen Zentrum für Hochschul- und Wissenschaftsforschung (DZHW) durchgeführt wird. Mit ihrer Teilnahme tragen Sie dazu bei, Wissenschaft besser zu verstehen.

Mit dieser Umfrage schauen wir besonders auf das Teilen und Nachnutzen von Forschungsdaten und Forschungssoftware. Ihre Erfahrungen und Meinungen sowie die Ihrer Kolleginnen und Kollegen bilden die empirische Grundlage für Empfehlungen an die Wissenschaftspolitik zur Verbesserung der Rahmenbedingungen für wissenschaftliche Arbeit.

Auch wenn Sie Forschungsdaten oder Forschungssoftware bisher noch nicht nutzen, freuen wir uns über Ihre Antworten.

Die Befragung dauert ca. 15 Minuten und umfasst ein breites Spektrum wichtiger Themen.

Zur Befragung gelangen Sie über den folgenden personalisierten Link:

[Survey-Link]

Sie können die Befragung jederzeit mithilfe Ihres Zugangscodes pausieren und fortsetzen: [PIN]

Für die Hochschul- und Wissenschaftsforschung ist Ihre Einschätzung von großem Wert. Wir würden uns daher sehr über Ihre Teilnahme freuen!

Mit herzlichem Dank für Ihre Unterstützung und freundlichen Grüßen,

Clemens Blümel und Judith Hartstein

Trendumfrage Forschungsdateninfrastrukturen

Deutsches Zentrum für Hochschul- und Wissenschaftsforschung, Berlin

Hinweise zum Datenschutz:

[Link zum Datenschutzblatt]

## 7.2 Informationsblatt zum Datenschutz

### Datenschutzblatt zur Trendumfrage „Forschungsdateninfrastrukturen“

#### Wer ist das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung (DZHW)?

Das DZHW ist eine gemeinnützige Einrichtung des Bundes und aller Länder und nutzt Befragungsdaten seit über vierzig Jahren für wissenschaftliche Untersuchungen zu bildungs- und hochschulpolitischen Fragestellungen. Ausführliche Informationen über uns und eine Reihe von Publikationen aus früheren Untersuchungen finden Sie unter [www.dzhw.eu](http://www.dzhw.eu).

Verantwortlich für die Studie im Sinne der DSGVO ist die DZHW GmbH, Lange Laube 12, in 30159 Hannover, vertreten durch die Geschäftsführerinnen Prof. Dr. Monika Jungbauer-Gans und Axel Tscherniak. Ansprechpartner gemäß Art. 26 DSGVO ist die DZHW GmbH. Projektverantwortliche für die Studie sind Clemens Blümel und Judith Hartstein.

#### Wie hat Sie die Einladung zur Teilnahme an der Befragung erreicht?

Sie bekommen diese Einladung, weil Sie als Professor:in oder wissenschaftliche/r Mitarbeiter:in an einer deutschen Universität beschäftigt sind. Die Namen, Titel und die Adressdaten aller Professor:innen und wissenschaftlichen Mitarbeiter:innen wurden auf den öffentlich zugänglichen Webseiten der Hochschulen recherchiert. Eine randomisiert ausgewählte Untergruppe dieser Personen wird zu dieser Befragung eingeladen.

#### Was passiert mit Ihren Daten?

- Ihre Angaben werden ausschließlich zu Forschungszwecken sowie forschungsunterstützenden Zwecken genutzt.
- Ihre Kontaktdaten nur für Einladungen zu Folgebefragungen.
- Externe Forscher:innen bzw. externes forschungsunterstützendes Personal erhalten nur anonymisierte Datensätze.
- Ihre Antworten werden getrennt von den Kontaktdaten zur Beantwortung verschiedenster Forschungsfragen analysiert.

Ebenso werden ausgewählte Paradata (Zeitpunkt und Dauer der Beantwortung der Befragung, Verweildauer auf einzelnen Fragebogenseiten, Art des verwendeten Endgeräts, Bildschirmauflösung) – in Verbindung mit den Befragungsdaten – ausschließlich für wissenschaftliche bzw. wissenschaftsunterstützende Zwecke ausgewertet.

Sie können der zukünftigen Verwendung Ihrer Kontaktdaten jederzeit widersprechen. In dem Fall werden Ihre Kontaktdaten gelöscht und nicht mehr für weitere Anschreiben verwendet. Die Rechtsgrundlage für diese Verarbeitungen ist Ihre uns gegenüber erklärte Einwilligung.

## Einwilligung

Ihre Teilnahme an der Befragung ist freiwillig. Durch die Nichtteilnahme entstehen Ihnen keinerlei Nachteile. Es ist selbstverständlich, dass alle gesetzlichen Bestimmungen des Datenschutzes eingehalten werden.

Wir versichern Ihnen, dass wir

- Ihre Kontaktdaten nicht mit Ihren Befragungs- und Parادات zusammen verarbeiten,
- Alle Befragungs- und Parادات ausschließlich zu Forschungszwecken und forschungsunterstützenden Zwecken nutzen und diese Wissenschaftler:innen und forschungsunterstützendem Personal außerhalb des DZHW nur in anonymisierter Form zur Verfügung stellen,
- Ihre Kontaktdaten bei Widerruf Ihrer Einwilligung sofort löschen

Weiterhin können Sie mit Blick auf Ihre personenbezogenen Daten gegenüber dem DZHW jederzeit die folgenden Rechte geltend machen

- Auskunftsrecht über Ihre gespeicherten Daten
- Recht auf Berichtigung Ihrer Daten
- Recht auf Einschränkung der Zwecke, für die die Daten genutzt werden,
- Recht auf Löschung
- Recht auf Widerruf Ihrer Einwilligung mit Wirkung für die Zukunft, sowie
- Recht auf Beschwerde bei der zuständigen Aufsichtsbehörde (Der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit, Graurheindorfer Str. 153, 53117 Bonn, poststelle@bfdi.bund.de).

Die Einhaltung der Vorkehrungen zum Schutz Ihrer Angaben wird durch den Datenschutzbeauftragten des DZHW, Herrn Martin Fuchs, überwacht. Bei Fragen zum Datenschutz oder wenn Sie eines der oben genannten Rechte ausüben wollen, können Sie sich an ihn wenden. Sie erreichen ihn unter der Rufnummer +49 (0)511 450670-491 oder unter datenschutz@dzhw.eu.

Für Auskünfte über das Forschungsprojekt stehen Ihnen die wissenschaftlichen Mitarbeiter gerne zur Verfügung: Clemens Blümel, [bluemel@dzhw.eu](mailto:bluemel@dzhw.eu), Judith Hartstein, [hartstein@dzhw.eu](mailto:hartstein@dzhw.eu).

## 7.3 Anonymisierte Variablen

Tabelle 6: Anonymisierte Variablen je Zugangsweg

Variablenname	Variablenlabel	On-Site-SUF	Remote-Desktop-SUF	Download-SUF
pid	CaseID (randomisiert)	freigegeben	freigegeben	freigegeben
einrichtung	Einrichtungstyp	freigegeben	freigegeben	freigegeben
posi	Statusgruppe	freigegeben	freigegeben	freigegeben
dfg10	DFG-Fachsystematik der Wissenschaftsbereiche und Fachgebiete (2020–2024)	freigegeben	freigegeben	freigegeben
dfg5	DFG-Fachsystematik der Wissenschaftsbereiche und Fachgebiete (2020–2024)	freigegeben	freigegeben	freigegeben
wibef_wb8_akadlebenslauf_a1	Erster Hochschulabschluss (Master, Diplom, o.ä.): Jahr	freigegeben	<b>anonymisiert</b>	<b>anonymisiert</b>
wibef_wb8_akadlebenslauf_a1_g1	Erster Hochschulabschluss (Master, Diplom, o.ä.): Jahr (aggregiert)	freigegeben	freigegeben	<b>anonymisiert</b>
wibef_wb8_akadlebenslauf_b1_g1	Erster Hochschulabschluss (Master, Diplom, o.ä.): Land (aggregiert)	freigegeben	freigegeben	<b>anonymisiert</b>
wibef_wb8_akadlebenslauf_b1_g2	Erster Hochschulabschluss (Master, Diplom, o.ä.): Land (aggregiert)	freigegeben	freigegeben	freigegeben <sup>a</sup>
wibef_wb8_akadlebenslauf_a2	Erste Promotion: Jahr	freigegeben	<b>anonymisiert</b>	<b>anonymisiert</b>
wibef_wb8_akadlebenslauf_a2_g1	Erste Promotion: Jahr (aggregiert)	freigegeben	freigegeben	<b>anonymisiert</b>
wibef_wb8_akadlebenslauf_b2_g1	Erste Promotion: Land (aggregiert)	freigegeben	freigegeben	<b>anonymisiert</b>
wibef_wb8_akadlebenslauf_b2_g2	Erste Promotion: Land (aggregiert)	freigegeben	freigegeben	freigegeben <sup>a</sup>
wibef_wb8_akadlebenslauf_a3	Erstberufung Jahr	freigegeben	<b>anonymisiert</b>	<b>anonymisiert</b>
wibef_wb8_akadlebenslauf_a3_g1	Erstberufung Jahr (aggregiert)	freigegeben	freigegeben	<b>anonymisiert</b>
wibef_wb8_akadlebenslauf_b3_g1	Erstberufung Land (aggregiert)	freigegeben	freigegeben	<b>anonymisiert</b>
wibef_wb8_akadlebenslauf_b3_g2	Erstberufung Land (aggregiert)	freigegeben	freigegeben	freigegeben <sup>a</sup>
wibef_de1_geschlecht	Geschlecht	freigegeben	freigegeben	<b>anonymisiert</b>
wibef_de1_geschlecht_g1	Geschlecht (aggregiert)	freigegeben	freigegeben	freigegeben
wibef_de2_geburtsjahr	Geburtsjahr	freigegeben	<b>anonymisiert</b>	<b>anonymisiert</b>
wibef_de2_geburtsjahr_g1	Geburtsjahr (aggregiert)	freigegeben	freigegeben	<b>anonymisiert</b>
wibef_de2_geburtsjahr_g2	Geburtsjahr (aggregiert)	freigegeben	freigegeben	freigegeben

<sup>a</sup> Für (Junior-)Professor\*innen anonymisiert.

## 7.4 Dokumentation der Herkunft sekundär genutzter Fragen

Tabelle 7: Dokumentation der Herkunft sekundär genutzter Fragen

Frage- nummer	Original (O) / Modifiziert (M)	Zitation
2.2	O	Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (2025). DZHW-Wissenschaftsbefragung 2023. Datenerhebung: 2023. Hannover: FDZDZHW. <a href="https://doi.org/10.21249/DZHW:scs2023:1.0.1">https://doi.org/10.21249/DZHW:scs2023:1.0.1</a> . „Fragebogen zur DZHW-Wissenschaftsbefragung 2023 (Kernfragebogen) (deutsch)“. Frage wb1, Seite 11.
2.3	O	Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (2025). DZHW-Wissenschaftsbefragung 2023. Datenerhebung: 2023. Hannover: FDZDZHW. <a href="https://doi.org/10.21249/DZHW:scs2023:1.0.1">https://doi.org/10.21249/DZHW:scs2023:1.0.1</a> . „Fragebogen zur DZHW-Wissenschaftsbefragung 2023 (Kernfragebogen) (deutsch)“. Frage wb2, Seite 13.
2.4	O	Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (2025). DZHW-Wissenschaftsbefragung 2023. Datenerhebung: 2023. Hannover: FDZDZHW. <a href="https://doi.org/10.21249/DZHW:scs2023:1.0.1">https://doi.org/10.21249/DZHW:scs2023:1.0.1</a> . „Fragebogen zur DZHW-Wissenschaftsbefragung 2023 (Kernfragebogen) (deutsch)“. Frage wb9, Seite 20 ff.
2.5	O	Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (2025). DZHW-Wissenschaftsbefragung 2023. Datenerhebung: 2023. Hannover: FDZDZHW. <a href="https://doi.org/10.21249/DZHW:scs2023:1.0.1">https://doi.org/10.21249/DZHW:scs2023:1.0.1</a> . „Fragebogen zur DZHW-Wissenschaftsbefragung 2023 (Kernfragebogen) (deutsch)“. Frage fp1, Seite 41.
2.6	O	Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (o. J.). DZHW-Wissenschaftsbefragung 2023. Datenerhebung: 2023. Hannover: FDZDZHW. Unveröffentlicht.
6.2	O	Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (2025). DZHW-Wissenschaftsbefragung 2023. Datenerhebung: 2023. Hannover: FDZDZHW. <a href="https://doi.org/10.21249/DZHW:scs2023:1.0.1">https://doi.org/10.21249/DZHW:scs2023:1.0.1</a> . „Fragebogen zur DZHW-Wissenschaftsbefragung 2023 (Kernfragebogen) (deutsch)“. Frage wb8, Seite 19.
6.3	O	Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (2025). DZHW-Wissenschaftsbefragung 2023. Datenerhebung: 2023. Hannover: FDZDZHW. <a href="https://doi.org/10.21249/DZHW:scs2023:1.0.1">https://doi.org/10.21249/DZHW:scs2023:1.0.1</a> . „Fragebogen zur DZHW-Wissenschaftsbefragung 2023 (Kernfragebogen) (deutsch)“. Frage fb2, Seite 26.
6.4	O	Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (2025). DZHW-Wissenschaftsbefragung 2023. Datenerhebung: 2023. Hannover: FDZDZHW. <a href="https://doi.org/10.21249/DZHW:scs2023:1.0.1">https://doi.org/10.21249/DZHW:scs2023:1.0.1</a> . „Fragebogen zur DZHW-Wissenschaftsbefragung 2023 (Kernfragebogen) (deutsch)“. Frage fb9, Seite 32.
6.5	O	Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (2025). DZHW-Wissenschaftsbefragung 2023. Datenerhebung: 2023. Hannover: FDZDZHW. <a href="https://doi.org/10.21249/DZHW:scs2023:1.0.1">https://doi.org/10.21249/DZHW:scs2023:1.0.1</a> . „Fragebogen zur DZHW-Wissenschaftsbefragung 2023 (Kernfragebogen) (deutsch)“. Frage de1, Seite 78.
6.6	O	Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (2025). DZHW-Wissenschaftsbefragung 2023. Datenerhebung: 2023. Hannover: FDZDZHW. <a href="https://doi.org/10.21249/DZHW:scs2023:1.0.1">https://doi.org/10.21249/DZHW:scs2023:1.0.1</a> . „Fragebogen zur DZHW-Wissenschaftsbefragung 2023 (Kernfragebogen) (deutsch)“. Frage de2, Seite 78.