



# Bericht zum 1. NFDI4Health User Survey

## Status-quo-Erhebung 2021 NFDI4Health

Ansprechpartner:innen

Katja Restel (restel@ub.uni-koeln.de)

Ralf Depping (depping@ub.uni-koeln.de)

Universitäts- und Stadtbibliothek Köln

Dezernat Forschungs- und Publikationsunterstützung

NFDI4Health User Research

Universitätsstraße 33

50931 Köln

**DOI:** 10.4126/FRL01-006461780

### Lizenz

Dieses Werk wurde unter der Lizenz „Creative Commons  
Namensnennung 4.0 International“ (CC BY 4.0) veröffentlicht.  
Den rechtsverbindlichen Lizenzvertrag finden Sie unter  
<https://creativecommons.org/licenses/by/4.0/legalcode>



## Vorbemerkungen

Deutschland verfügt über eine Fülle gesundheitsbezogener Daten aus gut konzipierten Kohortenstudien und Surveillanceprogrammen sowie aus klinischen Studien.

Das auf Anraten der Gemeinsamen Wissenschaftskonferenz im Rahmen der NFDI-Initiative für den Aufbau einer Nationalen Forschungsdateninfrastruktur gegründete NFDI4Health-Konsortium ging im Oktober 2020 mit der ersten Förderrunde für zunächst 5 Jahre an den Start. Das Konsortium konzentriert sich auf im klinischen, epidemiologischen und Public Health Kontext erhobene, personenbezogene Daten. Die Erhebung und Analyse dieser Daten sind ein wesentlicher Bestandteil bei der Entwicklung neuer Therapien und/oder präventiver Maßnahmen eines fortschrittlichen Gesundheitswesens.

Das interdisziplinäre NFDI4Health-Forschungskonsortium besteht aus Instituten, die jahrelange Erfahrung und Kompetenzen in den Bereichen Datenmanagement, Datenanalyse und Methoden mitbringen. Darüber hinaus ist es gut mit der Forschungsgemeinschaft verknüpft.

Geplant ist, eine umfassende Sammlung deutscher Studiendaten aus den Bereichen Epidemiologie, Public Health und klinischen Studien aufzubauen. Ziel ist ein zentralisierter Datenkatalog mit Suchfunktionen und einer Analyse Tool-Box. Das Suchen, Finden und Zugreifen auf dezentrale Daten soll vereinfacht und unter Beachtung aller datenschutzrechtlich und ethischen Aspekte möglich gemacht werden. Infrastrukturen, Schulungen, Services und Werkzeuge werden in Kooperation und im Einklang mit den Bedarfen der Nutzenden erarbeitet und umgesetzt.

Um den Datenkatalog und die Dienste eng an den Bedarfen der Forschenden Community auszurichten, wird von Anfang an mit einem Methoden Mix gearbeitet. Der vorliegende Survey ist der erste Schritt, eine Einsicht in den Status Quo zu bekommen und zu eruieren, inwieweit Datenmanagement, Standards und FAIR bereits Thema sind.

## Inhaltsverzeichnis

Vorbemerkungen.....	2
Methodisches Vorgehen .....	4
Zusammenfassung: Die Persona des Scientific Data Sharers .....	5
Ergebnisse der Bedarfserhebung .....	7
Beschreibung der Stichprobe/ Allgemeines.....	7
Forschungsdatenmanagement .....	9
Forschungsdatenbereitstellung .....	19
Forschungsdatennutzung .....	27
Schulungsbedarf .....	31
NFDI4Health .....	34
Zusammenfassung und Fazit .....	38
Ausblick .....	40

## Methodisches Vorgehen

Der Survey entstand innerhalb der Task Area 4.6 User Evaluation und Research, an der sich auch Personen beteiligen, die in anderen Task Areas des Projekts tätig sind. Zunächst wurde innerhalb des NFDI4Health-Konsortiums um Themen und Input gebeten, auf deren Grundlage eine erste Rohfassung des Befragungsinstruments im Februar/März 2021 entstand, die durch mehrere Feedbackrunden innerhalb der Task Area 4.6 ging und dann Ende Mai 2021 dem Konsortium für zweieinhalb Wochen zur Kommentierung vorgelegt wurde. Der Fragebogen wurde im Anschluss unter Berücksichtigung des Feedbacks aus dem Konsortium fertiggestellt. Ende Mai wurde ein Pre-Test mit zehn Personen vorgenommen, die sowohl aus den Kreisen des NFDI4Health-Projekts als auch von extern kamen. Das dadurch erlangte konstruktive Feedback wurde umgesetzt und einige inhaltliche wie optische Änderungen am Fragebogen vorgenommen.

Der Fokus des Surveys liegt auf den Themen Forschungsdatenmanagement, Forschungsdatenbereitstellung, Forschungsdatennutzung, Schulungsbedarf und NFDI4Health. Im Anschluss an die Themenbereiche wurden allgemeine Daten zur Fachdisziplin, akademischem Grad etc. erhoben, wobei diese Angaben freiwillig waren. Vorrangig sollen mit dem Survey sowohl die angesprochen werden, die Daten erstellen bzw. erheben und teilen, als auch jene, die Daten nachnutzen. Zu diesen drei Perspektiven gibt es jeweils einen Fragenblock. Da es sich eher um unterschiedliche Rollen und Perspektiven, aber nicht unbedingt unterschiedliche Personengruppen handelt, können die Teilnehmenden natürlich auch zu allen drei Perspektiven Stellung nehmen.

Die insgesamt 63 Fragen richteten sich somit nur teilweise an alle Teilnehmenden. Es wurde mit Filtern gearbeitet, um jene, die jeweils auf den Gebieten „Daten selber erheben“, „Daten zur Verfügung stellen“ bzw. „Daten nachnutzen“ nicht tätig sind, nicht mit unnötigen Fragen zu ermüden. Die Fragen wurden überwiegend als geschlossene oder halboffene Fragen formuliert, meist waren Mehrfachantworten möglich, darüber hinaus die Option, die Antwort „sonstige“ im Freitext genauer zu spezifizieren.

Die Online-Umfrage wurde mit der browserbasierten Software LimeSurvey<sup>1</sup> erstellt und durchgeführt. Der Fragebogen stand der Zielgruppe vom 21.6.2021 bis zum 4.9.2021 – also ca. 2½ Monate – zur Verfügung, der zeitliche Aufwand der Beantwortung betrug ca. 15–20 Minuten. Die Befragung erfolgte vollständig anonym. Die gewonnenen Daten wurden unter Berücksichtigung der gesetzlichen Datenschutzbestimmungen (DSGVO) erfasst und vertraulich behandelt. Personenbezogene Daten wurden lediglich in statistisch zusammengefasster Form dargestellt.

Die Zielgruppe wurde über verschiedene Kanäle auf den Survey aufmerksam gemacht und zur Teilnahme aufgefordert. Im Vorfeld wurden Personen innerhalb des Konsortiums gebeten, den Link zum Fragebogen gemeinsam mit einem kurzen Text über die von ihnen bespielten Verteiler bzw. Mailinglisten an einschlägige Verbände und Fachgruppen usw. weiterzuleiten. Des Weiteren wurden die Pressestellen der am Projekt beteiligter Institute gebeten, den Link zum Survey zu twittern oder ihre Mailinglisten für relevante Forschende zur Verbreitung zu nutzen. Auch das Konsortium als Ganzes wurde gebeten, den Link an Personen bzw. Personengruppen weiterzuleiten, die als relevant eingestuft wurden. Von dieser Vorgehensweise wurde sich ein Schneeballeffekt innerhalb der Community der medizinisch Forschenden versprochen.

---

<sup>1</sup> <https://www.limesurvey.org/de/>

## Zusammenfassung: Die Persona des Scientific Data Sharers

Um die umfang- und detailreichen Ergebnisse des Surveys prägnant zusammenzufassen, wird die fiktive Persona *SciDaS* (Scientific Data-Sharer) modelliert. *SciDaS* steht an dieser Stelle stellvertretend für ein typisches Antwortverhalten, auf eine Differenzierung des Antwortverhaltens wird also an dieser Stelle bewusst verzichtet. Die angegebenen Werte beziehen sich jeweils auf die für die Frage ermittelte Grundgesamtheit, die je nach Frage zwischen 222 und 108 schwankt.

### Allgemeines

*SciDaS* ist promoviert und an einer deutschen außeruniversitären Forschungseinrichtung oder Hochschule im Bereich Epidemiologie, Public Health oder<sup>2</sup> Forschung tätig. An der jeweiligen Institution erhebt, verwaltet, analysiert *SciDaS* Daten oder ist im Projektmanagement, in der Koordination oder Leitung tätig. Die Daten, die *SciDaS* behandelt, gehören zum Feld der bevölkerungsbasierten Umfrage oder stammen aus dem Gesundheitswesen.

### Forschungsdatenmanagement

*SciDaS* erstellt oder verwaltet überwiegend personenbezogene Daten. Die meisten davon fallen unter die Kategorien „strukturierte numerische Daten“, „Interviewdaten“ oder „unstrukturierte nicht-numerische Daten“. Die Daten werden mit Hilfe von Fragebögen, Interview-Transkripten oder biochemischen Verfahren gewonnen.

Die jeweiligen Institutionen verwenden zur Formatierung oder Sicherung keine Standards bzw. zumindest sind sie *SciDaS* nicht bekannt. Für die Erfassung der Daten hingegen wird überwiegend die International Classification of Diseases (ICD) verwendet.

*SciDaS* benutzt keine externen Systeme für die Forschungsdatenverwaltung. Die Daten werden während des Forschungsprozesses auf dem Institutsserver oder dem Dienstrechner, danach auf dem Instituts- oder dem Universitätsserver gesichert. Mit passenden Repositorien kennt sich *SciDaS* nicht aus.

### Forschungsdatenbereitstellung

*SciDaS* ist durchaus teilfreudig, es werden aufbereitete Daten oder Erhebungsinstrumente mit Forschenden aus der eigenen Institution oder Forschenden, die persönlich bekannt sind, geteilt. Dies geschieht primär per E-Mail, über persönliche Kommunikationswege oder über Cloud-Server (z.B. Dropbox). Die Motivation hinter dem Teilen von Daten besteht in dem Wunsch, Kooperationen zu fördern, Validierung und Verifizierung weiterzutreiben, das Forschungsgebiet voranzubringen und nicht zuletzt darin, dass andere auch teilen.

Generell würde *SciDaS* auch Metadaten, Instrumente oder Studiendokumente mit anderen teilen, allerdings vorzugsweise mit Forschenden, die persönlich bekannt sind. Dies sollte dann auf Anfrage geschehen oder unter der Voraussetzung, dass eine konkrete Nutzenabsprache getroffen wird, die Daten formal zitiert werden oder der:die Datengebende als Co-Autor:in benannt wird.

Es besteht zudem die Bereitschaft, Daten für Studienteilnehmende in aggregierter Form zur Verfügung zu stellen.

Bei der Bereitstellung der Daten achtet *SciDaS* nach eigenen Angaben auf Maschinenverarbeitbarkeit.

---

<sup>2</sup> Da bei den meisten Fragen Mehrfachantworten möglich waren, müsste es „und/oder“ bei den Aufzählungen heißen, weil dies aber den Lesefluss stört, wird im Folgenden „oder“ verwendet.

## Forschungsdatennachnutzung

*SciDaS* nutzt eher selten Daten nach, dann aber, um die eigene Fragestellung zu schärfen oder einen Überblick zu erhalten. *SciDaS* sucht gezielt nach strukturierten numerischen Daten, hat aber prinzipiell Interesse an allen Forschungsmaterialien. Sie werden gezielt auf Projekt-Webseiten gesucht, es bestehen also schon weiterführende Kenntnisse bei der Suche nach Daten auf dezentralen Seiten. Bei der Suche nach Daten zur Nachnutzung begegnen *SciDaS* einige Störfaktoren; die drei gravierendsten sind hier (1) ungenügende Beschreibung/Erschließung der Daten, (2) komplizierte Prozedere der Datenbeschaffung und (3) fehlende Interoperabilität.

## Schulungen

*SciDaS* benötigt Schulungen in den meisten genannten Bereichen auf Expertenlevel, außer bei den Themen „Daten teilen und publizieren“, „Daten nachnutzen“ und „Daten archivieren“ – in diesen Bereichen auch auf Basis-/Einsteigerlevel. Zu diesen Ergebnissen sei an dieser Stelle angemerkt, dass der Teil der Forschenden-Community, der Kenntnis von der Umfrage hatte bzw. daran teilnahm (wie bereits im vorherigen Abschnitt erwähnt), NFDI4Health nahesteht und demnach in der Forschungsdatenmanagement (FDM-)Thematik deutlich sensibilisierter und auch erfahrener ist. Demnach kann angenommen werden, dass die Bedarfe innerhalb der nicht durch den Survey erreichten Forschenden größer sind und unter Umständen auch eher Formate auf niedrigerem Niveau nachgefragt sind.

*SciDaS* möchte seinen Informationsbedarf am liebsten über (Online-)Seminare decken.

An der Institution, an der *SciDaS* tätig ist, gibt es keine zentrale Anlaufstelle für Forschungsdatenmanagement.

## NFDI4Health

Wie bereits im Forschungsdesign und auch im letzten Absatz erwähnt, wurde der Link zum Survey über Multiplikator:innen aus dem NFDI4Health-Konsortium gestreut, was dazu führte, dass viele der erreichten Personen nicht nur schon ein Bewusstsein für FDM hatten, sondern auch NFDI4Health kannten. Deshalb ist es auch wenig verwunderlich, dass unsere Persona *SciDaS* NFDI4Health kennt, weil sie:er entweder selber am Projekt angebunden ist oder ihm:ihr dieses aus der eigenen Institution bekannt ist.

*SciDaS* stellt demnach auch die gleichen Anforderungen an eine föderale Dateninfrastruktur wie die im Antrag identifizierten und beurteilt alle von NFDI4Health avisierten Serviceleistungen als wichtig bis eher wichtig.

### **Konkrete Handlungsbedarfe:**

- **Etablierung von einheitlichen Sicherheitsstandards**
- **Erhöhung der Awareness für Interoperabilität der Daten**
- **gesteigertes Bewusstsein dafür schaffen, dass Forschungsdaten auf überregionalen Repositorien zur Nachnutzung gespeichert werden sollten**
- **Schulungen und Guidelines**
- **rechtliche Grundlagen, bindende Musterformulare für Nutzungsabsprachen**
- **zentrale Plattform zur gezielten Suche für Daten zur Nachnutzung**

## Ergebnisse der Bedarfserhebung

### Beschreibung der Stichprobe/ Allgemeines

Im Befragungszeitraum vom 21.6.2021 bis 4.9.2021 haben insgesamt 339 Personen die Startseite der Umfrage aufgerufen. 222 Personen haben die Umfrage gestartet, davon haben 49 % (108 Teilnehmende) den Fragebogen vollständig ausgefüllt.

Durch die gezielte Verbreitung der Umfrage über Multiplikator:innen aus dem NFDI4Health-Konsortium, aber auch durch das gezielte Anschreiben von Pressestellen der relevanten Institutionen, konnte die Zielgruppe erfolgreich adressiert werden. Der Survey erreichte vornehmlich in Deutschland arbeitende (98 %) Datennutzende bzw. Datenbereitstellende an Hochschulen (39 %) und außeruniversitären Forschungseinrichtungen (44 %).



Abbildung 1 „Wo liegt Ihr derzeitiger Arbeitsort?“ (n=108).

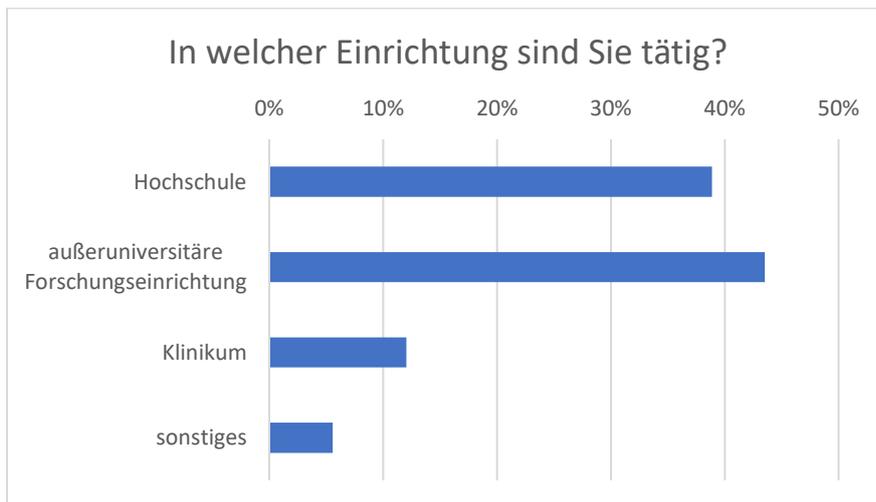


Abbildung 2 „In welcher Einrichtung sind Sie tätig?“ (n=108).

Erwartungsgemäß wurden zudem überwiegend Personen mit Promotion (58%) oder Habilitation (20%) erreicht.

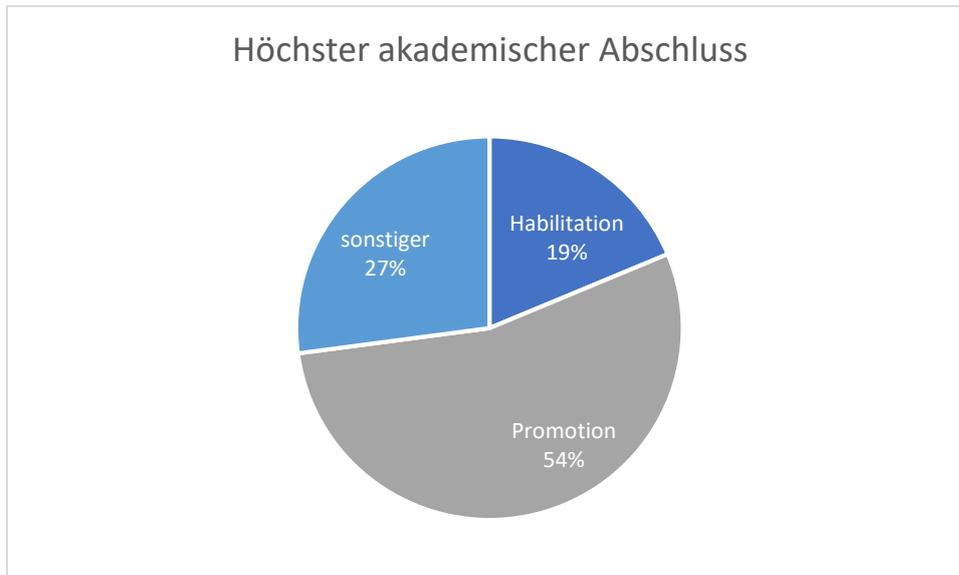


Abbildung 3 „Welchen höchsten akademischen Grad haben Sie?“ (n=107).

Bei der Frage nach den Tätigkeitsbereichen, bei der sich die Antwortkategorien auf die einzelnen medizinischen Bereiche fokussierten, liegt ein klarer Schwerpunkt bei Epidemiologie und Medizinischer Biometrie (53 Nennungen bei n=107), Forschung (44 Nennungen) und Public Health (37 Nennungen).

Die meisten Befragten befassen sich im beruflichen Kontext mit der Erhebung, Verwaltung oder Analyse von Daten (80 Nennungen bei n=107) und/oder mit Projektmanagement, Koordination und Leitung (62 Nennungen). Dies zeigt, dass wir mit dem Bogen die Daten-Gebenden, aber auch die potentiellen Daten-Nehmenden erreicht haben.

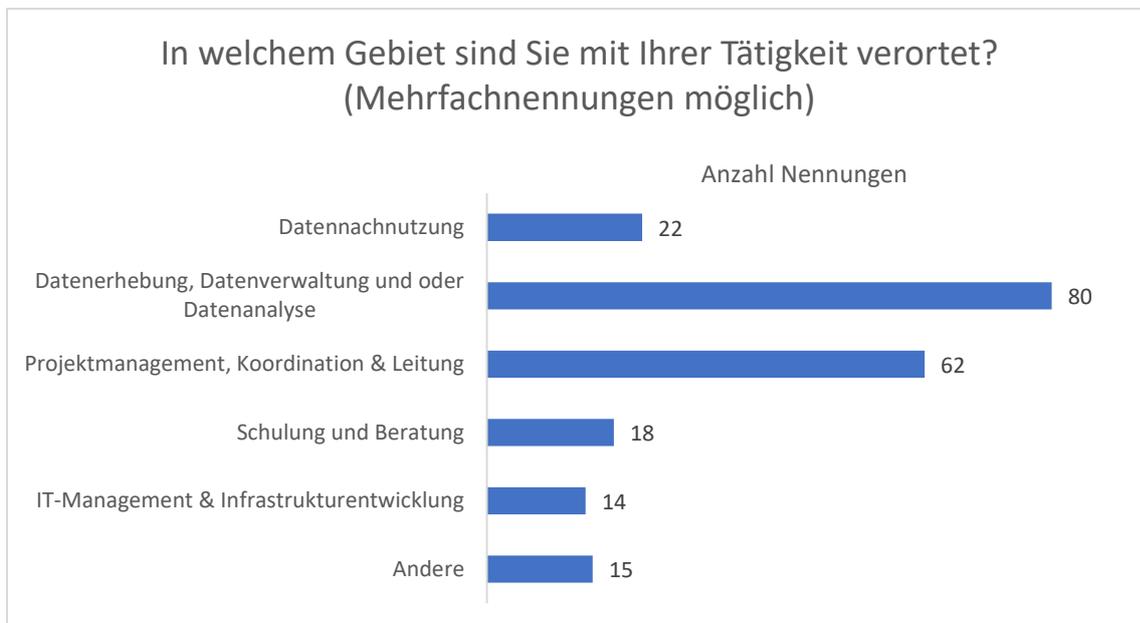


Abbildung 4 „In welchem Gebiet sind Sie mit Ihrer Tätigkeit verortet?“. Mehrfachnennungen möglich (n=107, Anzahl Nennungen 211).

Die Daten, die von den Befragten erhoben, verwaltet und analysiert werden, stammen primär aus bevölkerungsbasierten Umfragen (64 Nennungen bei n=107), klinischen Studien (50 Nennungen) und/oder aus dem Gesundheitswesen (50 Nennungen).

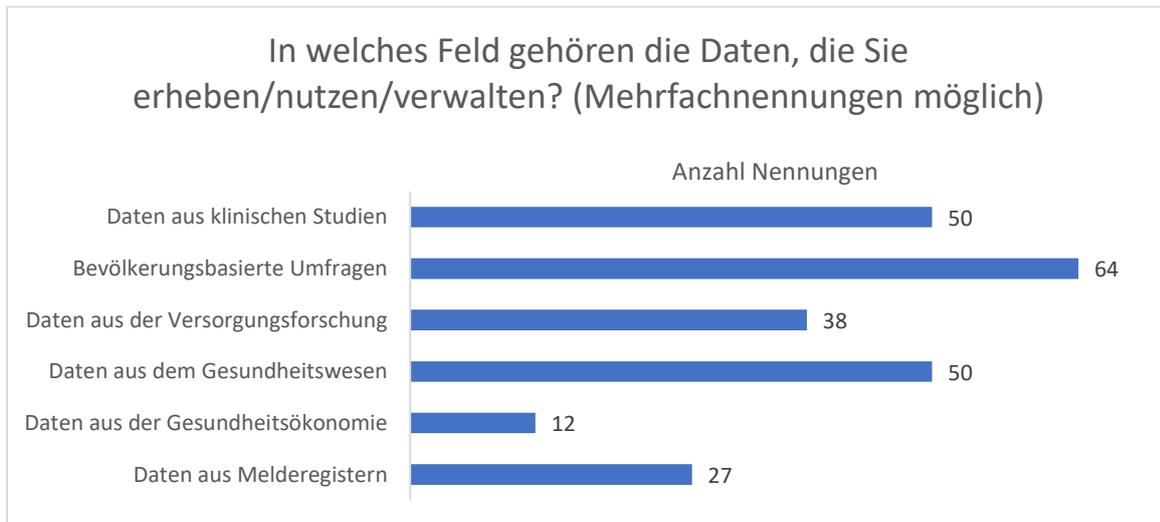


Abbildung 5 „In welches Feld gehören die Daten, die Sie erheben/nutzen/verwalten?“. Mehrfachnennungen möglich (n=107, Anzahl Nennungen 241).

## Zusammenfassung

Anhand der bisher beschriebenen Daten aus dem Teil *Allgemeines* ist ersichtlich, dass die Umfrage überwiegend Personen erreicht hat, die mit ihr auch erreicht werden sollten: Forschende in Deutschland, die sich mit medizinischen Forschungsdaten aus den wichtigsten Kerngebieten befassen. Die Teilnahme an der Befragung war natürlich freiwillig, somit ist die Zusammensetzung der Teilnehmenden selbstselektiv entstanden. Es gibt gute Gründe für die Annahme, dass Personen, die sich an der Befragung beteiligt haben, tendenziell eher einen stärkeren Bezug zum Thema Forschungsdatenmanagement haben als die Grundgesamtheit aller medizinisch Forschenden.

## Forschungsdatenmanagement

Im Rahmen des Fragenblocks *Forschungsdatenmanagement* wurden direkt zu Beginn Filter verwendet, um jene Befragten weiterzuleiten, die weder selber Forschungsdaten erheben noch Forschungsdaten verwalten, die von anderen erhoben wurden. Dies traf auf 24 % der 225 Befragten zu, die diese Fragen beantworteten. Diese übersprangen somit den Teil mit den Fragen zum Forschungsdatenmanagement und wurden direkt zum nächsten Themenkomplex „Forschungsdatenbereitstellung“ weitergeleitet.

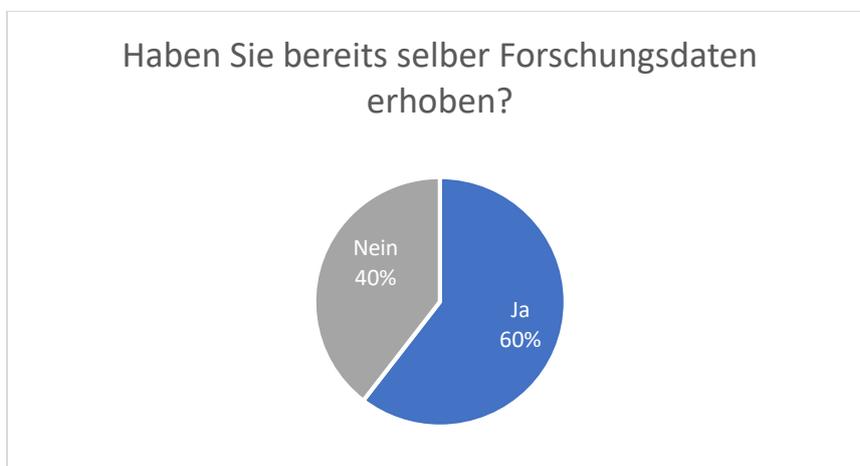


Abbildung 6 „Haben Sie bereits selber Forschungsdaten erhoben?“ (n=222)

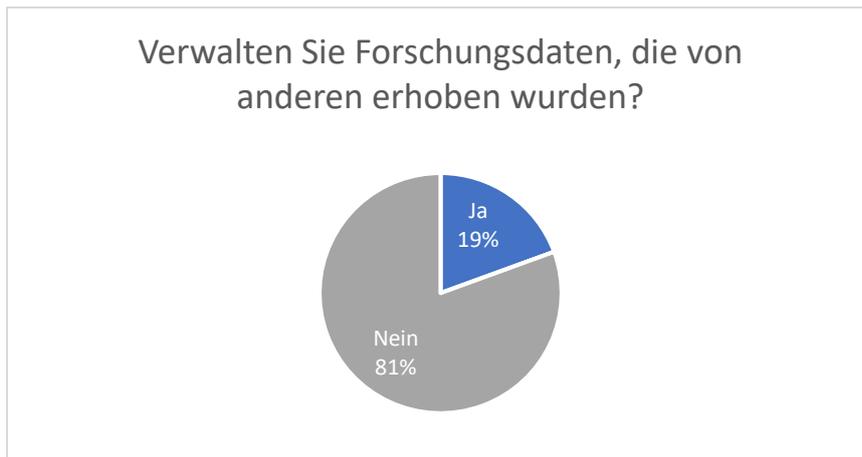


Abbildung 7 „Verwalten Sie Forschungsdaten, die von anderen erhoben wurden?“ (n=67)

Der Teil *Forschungsdatenmanagement* enthält Fragen zur Art und Strukturierung der von den Befragten erhobenen bzw. verwalteten Daten, aber auch nach in den jeweiligen Institutionen der Befragten angewandten Standards. Zudem werden Fragen zu verwendeten Tools, Interoperabilität und Speicherorten gestellt.

Etwas weniger als die Hälfte der Befragten hat bereits einmal einen Datenmanagementplan erstellt (48%).

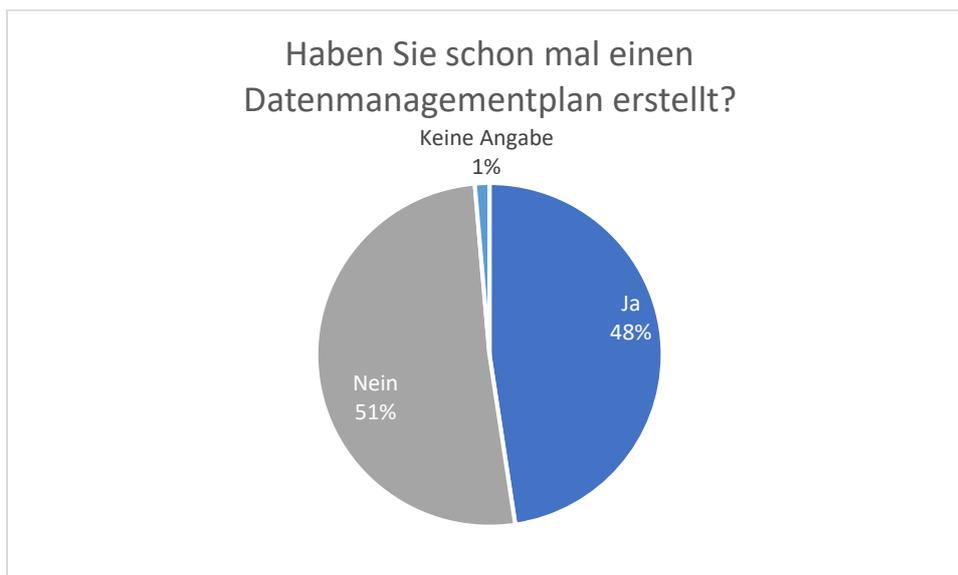


Abbildung 8 „Haben Sie schon mal einen Datenmanagementplan erstellt?“ (n=147)

Wenn ein Datenmanagementplan angefertigt wurde, geschah dies, weil es Fördervorgaben entsprach (31 Nennungen bei n=70) und/oder aus Eigeninteresse (35 Nennungen). Unter „Sonstiges“ wurde zudem noch Datenschutz (5 von n=12) als Grund genannt.

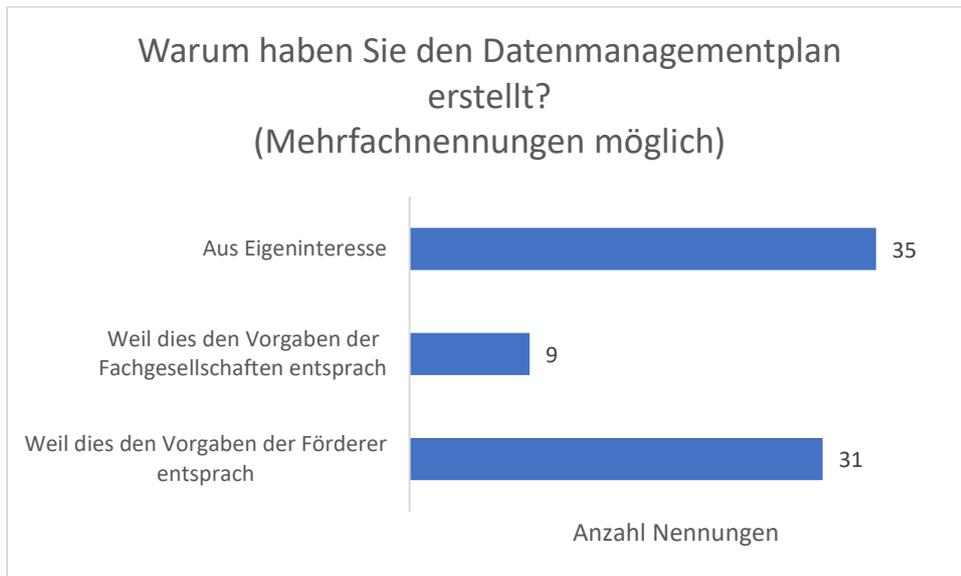


Abbildung 9 „Warum haben Sie den Datenmanagementplan erstellt?“ Mehrfachnennungen möglich (n=70, Anzahl Nennungen 75)

Die Vertretenden der befragten Community erheben bzw. erstellen am häufigsten strukturierte numerische Daten (134 Nennungen bei n=145), Interviewdaten (95 Nennungen) und/oder strukturierte nicht-numerische Daten (76 Nennungen).

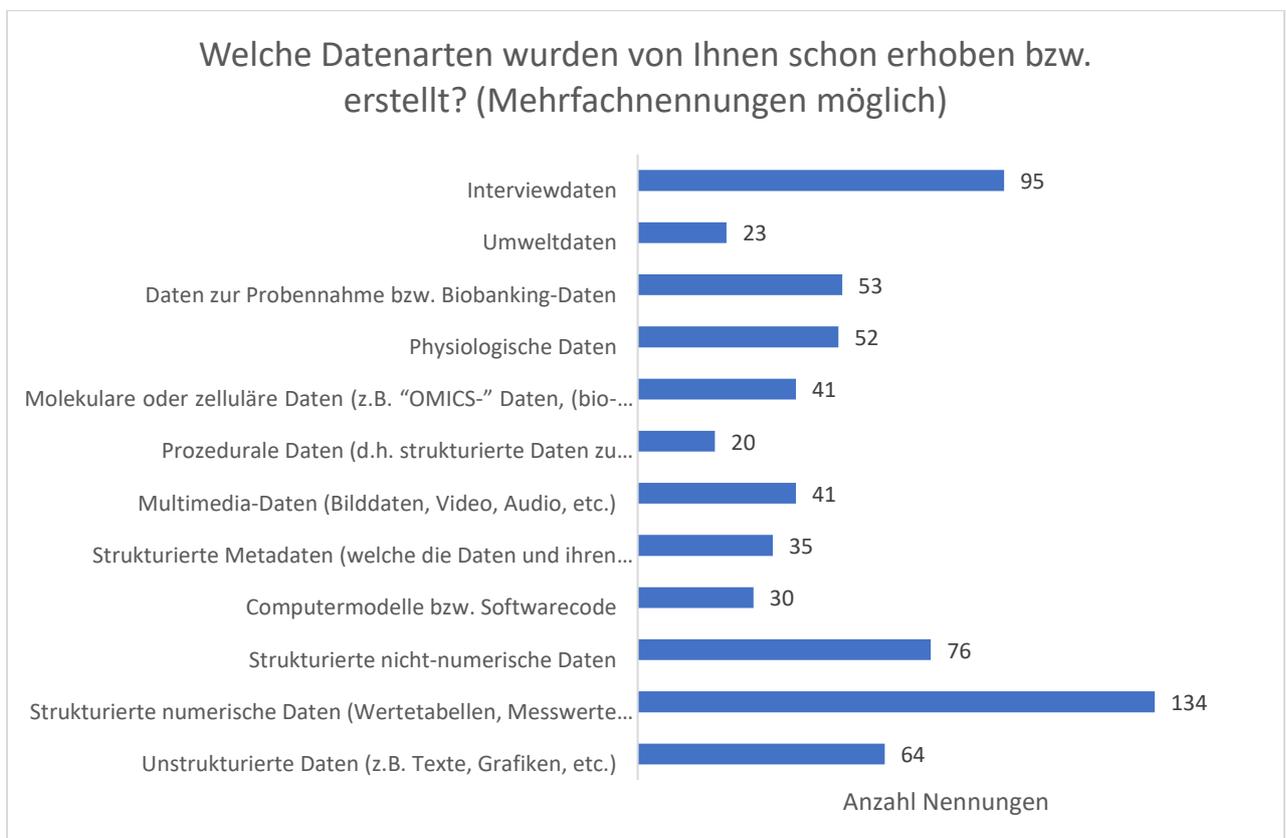


Abbildung 10 „Welche Datenarten wurden von Ihnen schon erhoben bzw. erstellt?“ Mehrfachnennungen möglich (n=145, Anzahl Nennungen 664)

Gut drei Viertel der Befragten erheben personenbezogene Daten.

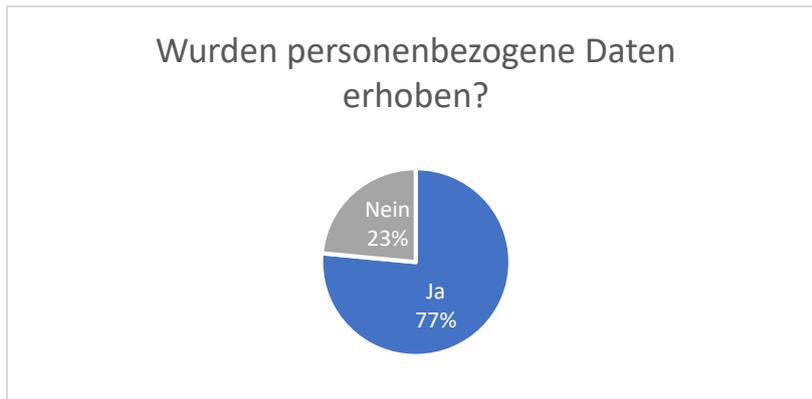


Abbildung 11 „Wurden personenbezogene Daten erhoben?“ (n=149)

Zur Datenerhebung verwenden die meisten Befragten Fragebögen (118 Nennungen bei n=149) oder (zusätzlich) Interview-Transkripte (62 Nennungen). Des Weiteren spielen biochemische Verfahren (47 Nennungen) und Exporte aus Primärsystemen (51 Nennungen) eine Rolle.

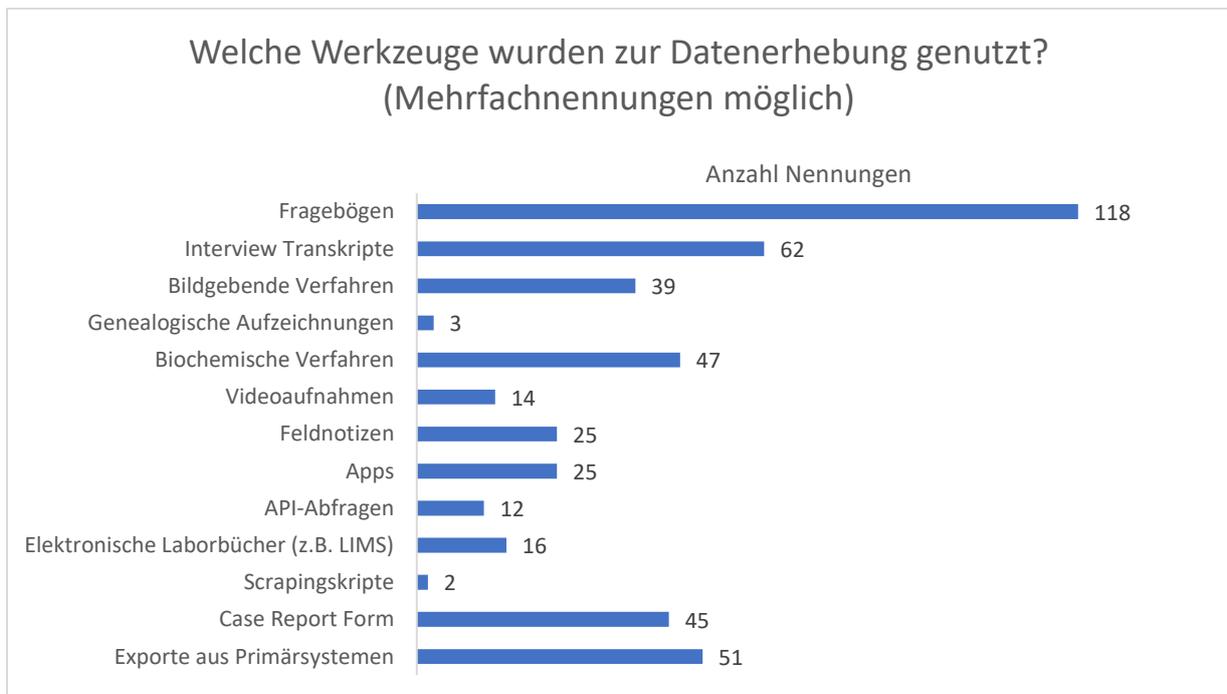


Abbildung 12 „Welche Werkzeuge wurden zur Datenerhebung genutzt?“ Mehrfachnennungen möglich (n=149, Anzahl Nennungen 459)

## Standards

Der nächste Abschnitt befasst sich mit den Ergebnissen der Abfrage der an den Institutionen bzw. in den Projekten verwendeten Standards zur Sicherung, Formatierung, Beschreibung, Erfassung und Dokumentation von Forschungsdaten.

41 % der Befragten wissen nicht, ob an ihrer Institution bzw. in ihren Projekten Vorgaben oder Standards zur Sicherung von Daten existieren.

Von den 44 %, die mit *ja* geantwortet haben, gaben 35 Personen an, dass (interne) Standard Operating Procedures (SOPs) hierzu verwendet würden oder sich nach den Vorgaben der IT bzw. des Rechenzentrums der Institution gerichtet würde. Der Fokus bei den Sicherungsstandards liegt bei den internen Regelungen der Institute bzw. Einrichtungen.

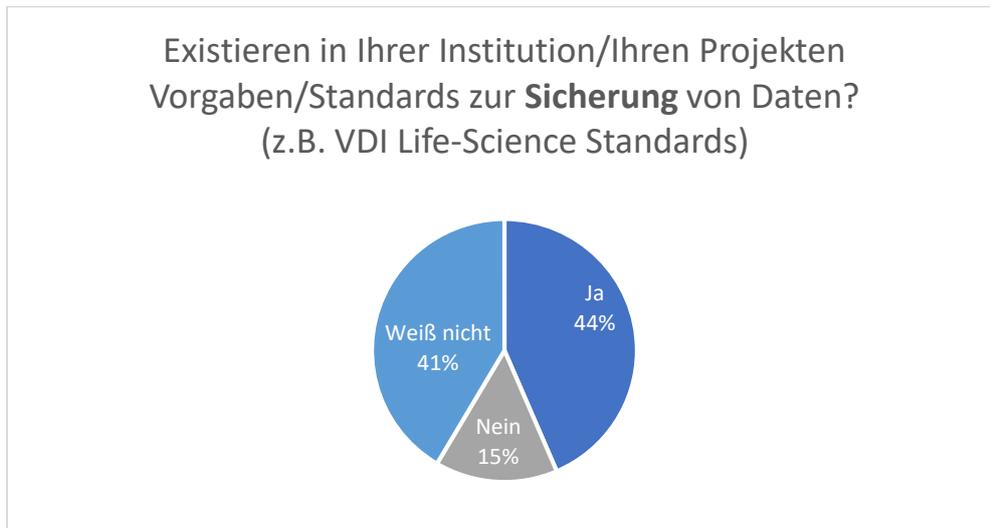


Abbildung 13 „Existieren in Ihrer Institution/Ihren Projekten Vorgaben/Standards zur **Sicherung** von Daten? (z.B. VDI Life-Science Standards)“ (n=148)

Bei der Frage nach den Standards zur Formatierung der erfassten oder zu erfassenden Daten, äußern 39 % der Befragten Unwissen und 36 % gaben an, solche Standards gebe es nicht. Bei denjenigen, die die Frage positiv oder mit *teilweise* beantworteten (25 % der Befragten), gibt es 16 konkrete Nennungen, wobei Clinical Data Interchange Standards (CDISC) und Operational Data Model (ODM) mit jeweils drei Nennungen am häufigsten erwähnt werden.

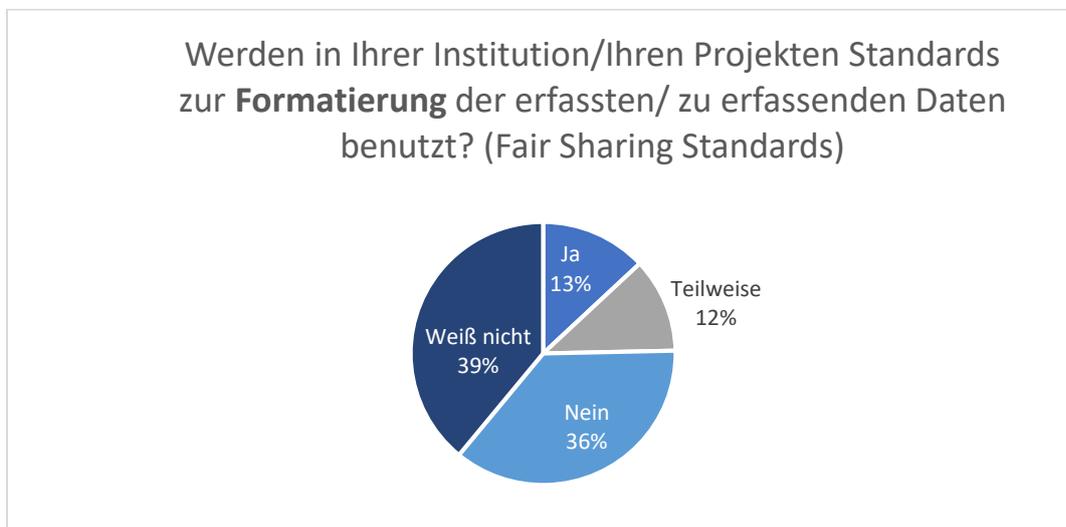


Abbildung 14 „Existieren in Ihrer Institution/Ihren Projekten Vorgaben/Standards zur **Formatierung** der erfassten/zu erfassenden Daten? (Fair Sharing Standards)“ (n=145)

Mehr als ein Drittel (37 %) der Befragten weiß nicht, ob an ihrer Institution bzw. in ihrem Projekt Standards zur Beschreibung der Daten verwendet werden, ca. ein Drittel gibt an, dass dies nicht so sei. Diejenigen, die die Fragen bejahen oder mit *teilweise* beantworten (jeweils ca. 15 %), nennen hierfür sehr heterogene Standards von internen Lösungen bis CDISC.

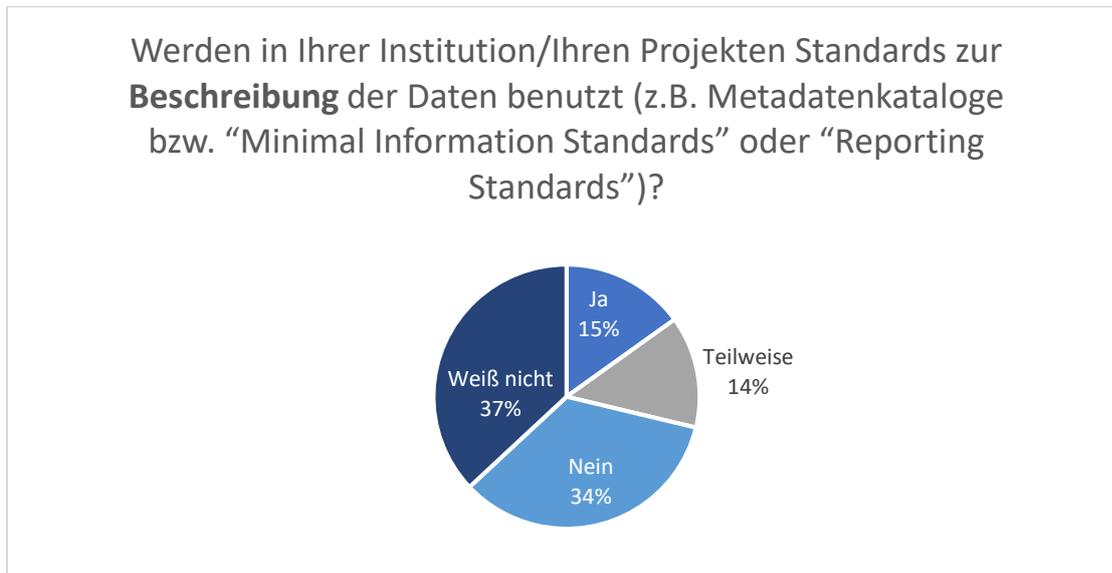


Abbildung 15 „Werden in Ihrer Institution/Ihren Projekten Standards zur **Beschreibung** der Daten benutzt (z.B. Metadatenkataloge bzw. „Minimal Information Standards“ oder „Reporting Standards“)?“ (n=145)

Bei der Frage nach der Verwendung von Standardterminologien bei der Erfassung von Daten antwortet ein Drittel mit *ja* (33 %) und mehr als ein Viertel mit *teilweise* (27 %). Auch hier liegt der Anteil derjenigen, die darüber nichts wissen, bei über einem Viertel (27 %). Hier wird bei denjenigen, die nicht mit *nein* oder *weiß nicht* antworten (61 % der Befragten), zum Großteil ICD (20 Nennungen von n=37) als Standard angegeben.

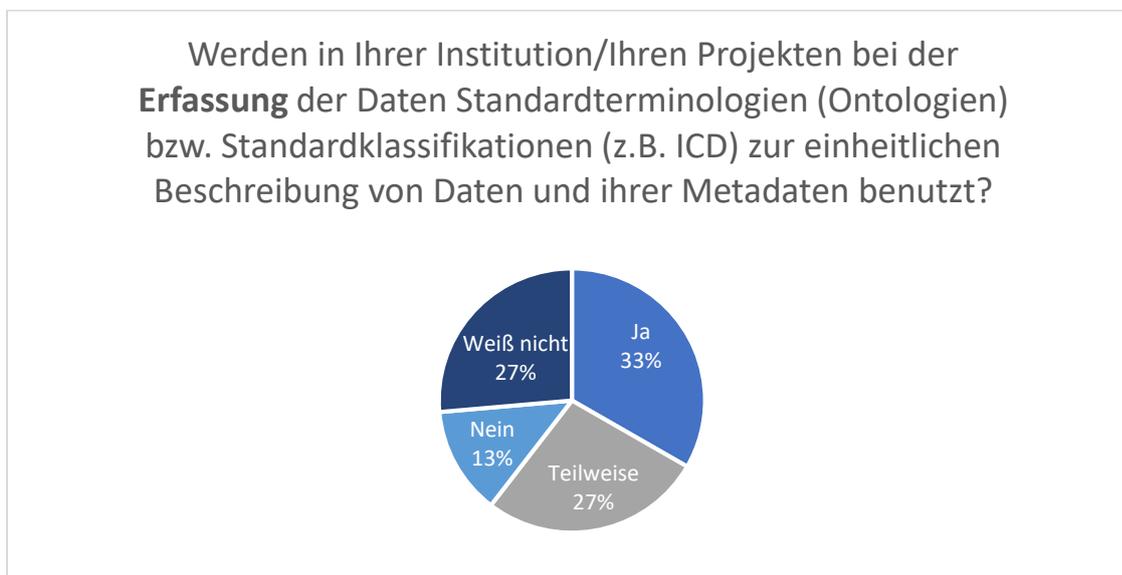


Abbildung 16 „Werden in Ihrer Institution/Ihren Projekten bei der Erfassung der Daten Standardterminologien (Ontologien) bzw. Standardklassifikationen (z.B. ICD) zur einheitlichen Beschreibung von Daten und ihrer Metadaten benutzt?“ (n=143)

Insgesamt zeichnet sich hinsichtlich der Antworten zu den Fragen nach existierenden Standards eine Tendenz zu *weiß nicht* aus. Auch, dass häufig Verneinungen vorkommen oder interne Lösungen genannt werden, lässt darauf schließen, dass die Verwendung allgemeiner Standards noch nicht vorherrscht.

Auch bei der verwendeten Software bestehen kaum Gemeinsamkeiten innerhalb der befragten Stichprobe. Lediglich *ArXiv* bekommt 18 Nennungen bei  $n=132$ . Neben den vorgegebenen Programmen nennen 25 Befragte weitere Programme und Software, wobei hier eine Tendenz zu individuellen internen Lösungen sichtbar wird.

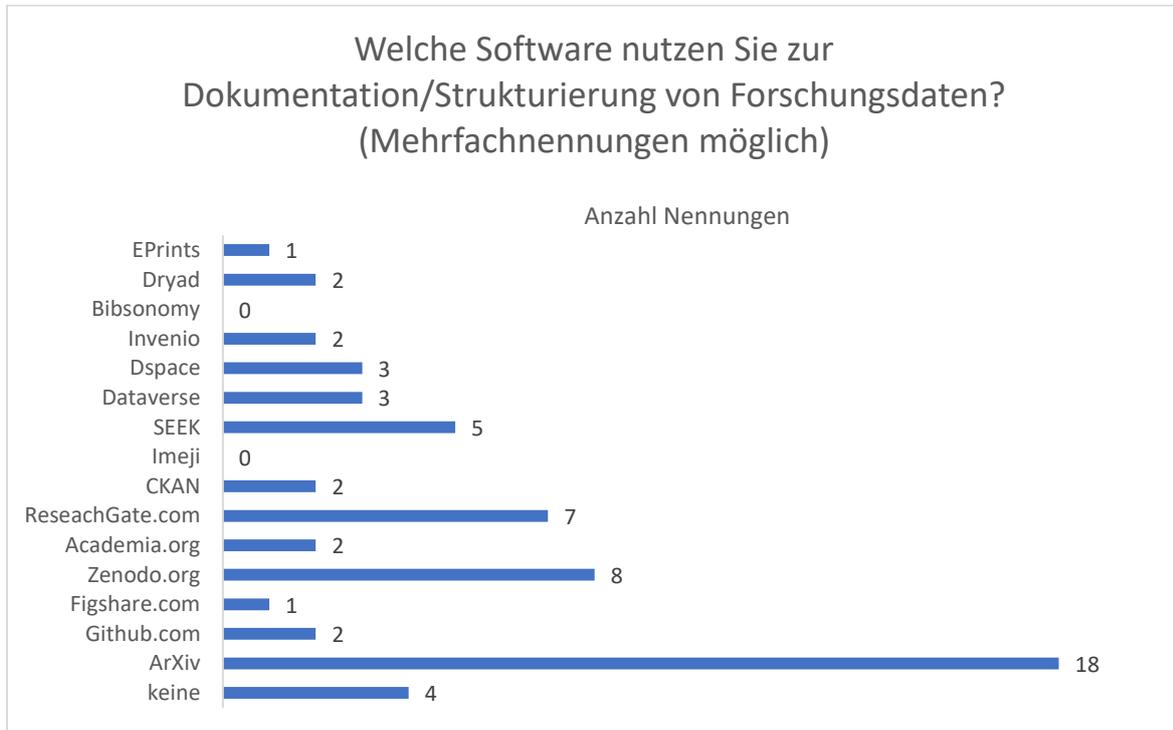


Abbildung 17 „Welche Software nutzen Sie zur Dokumentation/Strukturierung von Forschungsdaten?“ (Mehrfachnennungen möglich) ( $n=50$ ; 60 Nennungen)

Die nächste Frage befasste sich mit der Interoperabilität zwischen den von den Befragten verwendeten Systemen zum Forschungsdatenmanagement und extern verwendeter Software, Diensten oder Systemen. Auf diese Frage antworteten mehr als die Hälfte der Befragten (52 %), dass sie keine externe FDM-Software, -Systeme oder -Dienste nutzen, und ca. ein Fünftel (21 %) kann diese Frage nicht beantworten.

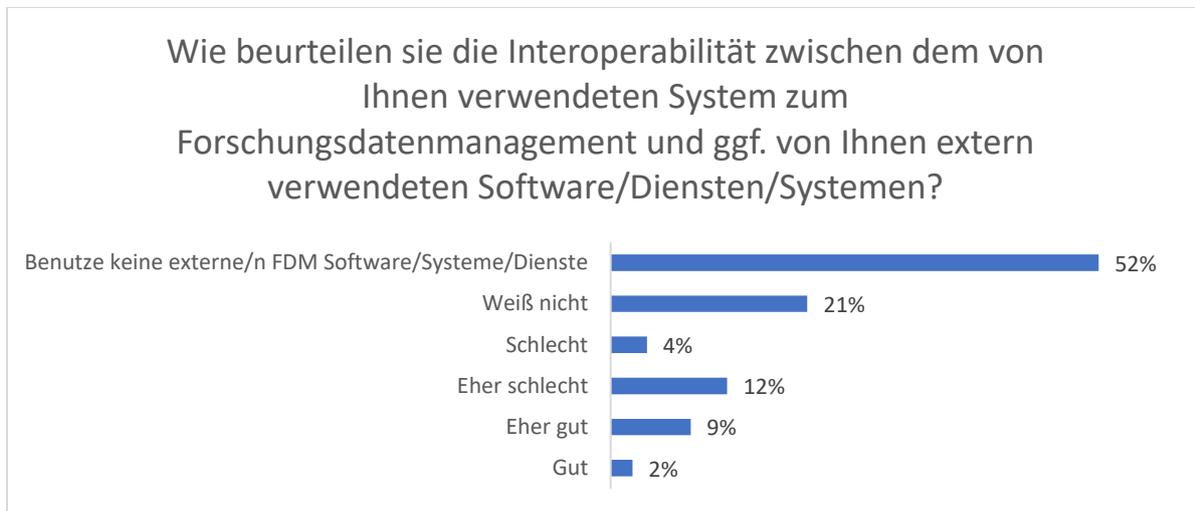


Abbildung 18 „Wie beurteilen sie die Interoperabilität zwischen dem von Ihnen verwendeten System zum Forschungsdatenmanagement und ggf. von Ihnen extern verwendeten Software/Diensten/Systemen?“ (n=129)

Dies schlägt sich auch in den Ergebnissen der anschließenden Frage zur Bereitschaft, Metadaten auf einer Metadatenplattform bereitzustellen, die ein anderes System bzw. Format verwendet, nieder: Hier antworten 43 % mit *weiß nicht*. Allerdings stimmen insgesamt ebenfalls 43 % der Befragten der Aussage zu (12 %) bis eher zu (31 %), dass die Bereitschaft zu Teilen sinke, wenn ein anderes System bzw. Format verwendet wird, woraus sich ableiten lässt, dass fremde Systeme oder Formate ein Hindernis zur Metadatenbereitstellung darstellen können.

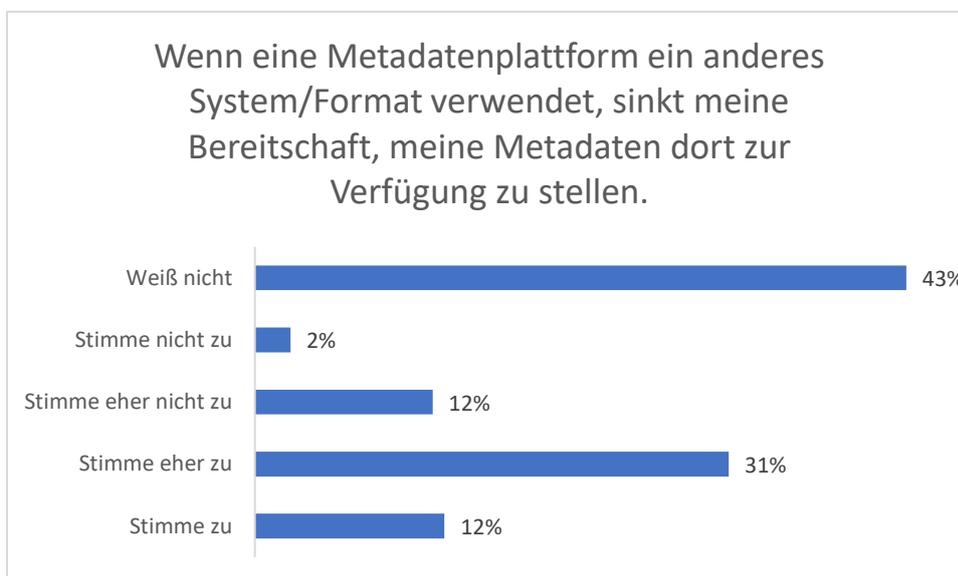


Abbildung 19 „Wenn eine Metadatenplattform ein anderes System/Format verwendet, sinkt meine Bereitschaft, meine Metadaten dort zur Verfügung zu stellen.“ (n=129)

Die Befragten speichern ihre Daten primär in proprietären Formaten (88 Nennungen bei n=129).

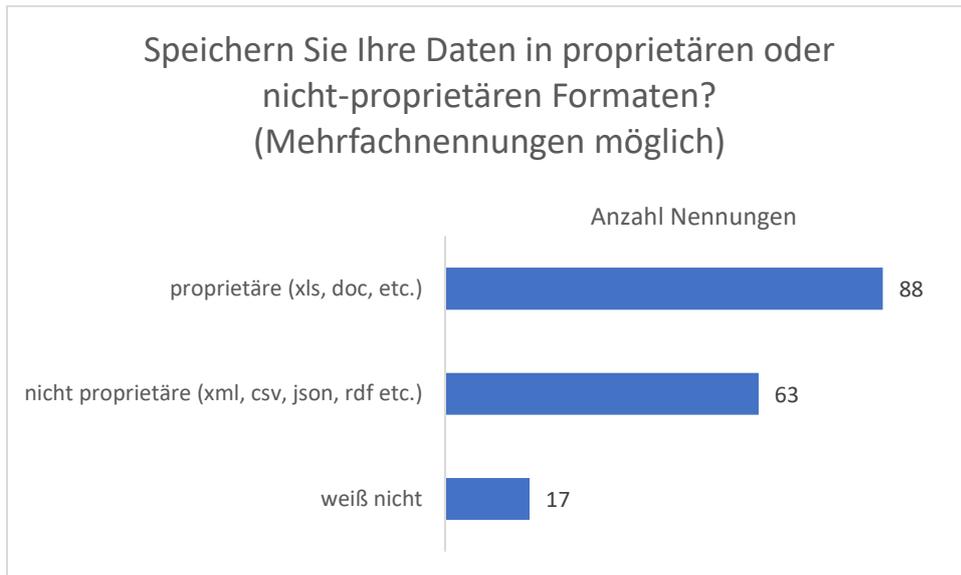


Abbildung 20 „Speichern Sie Ihre Daten in proprietären oder nicht-proprietären Formaten?“ Mehrfachnennungen möglich (n=129, Anzahl Nennungen 168)

Während des Forschungsprozesses werden die Daten überwiegend auf dem Institutsserver (102 Nennungen bei n=129), aber auch auf dem Dienstrechner abgelegt.

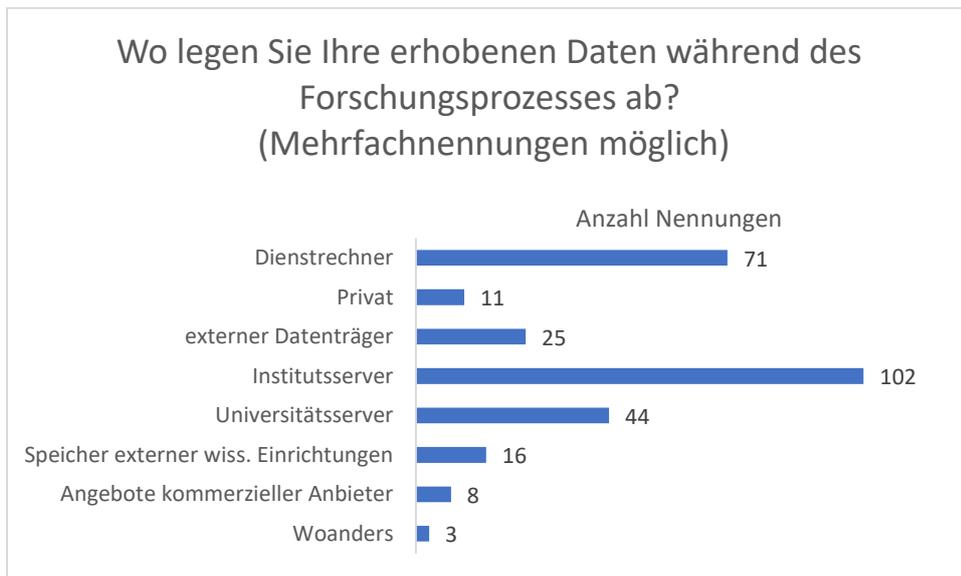


Abbildung 21 „Wo legen Sie Ihre erhobenen Daten während des Forschungsprozesses ab?“ Mehrfachnennungen möglich (n=129, Anzahl Nennungen 280)

Ein ähnliches Bild zeigt sich bei der Frage nach dem Speicherort der Daten, nachdem der Forschungsprozess abgeschlossen wurde. 94 von 129 Befragten verweisen auf den Institutsserver, 44 (auch) auf den Universitätsserver. Die Speicherung von Forschungsdaten erfolgt damit überwiegend lokal.

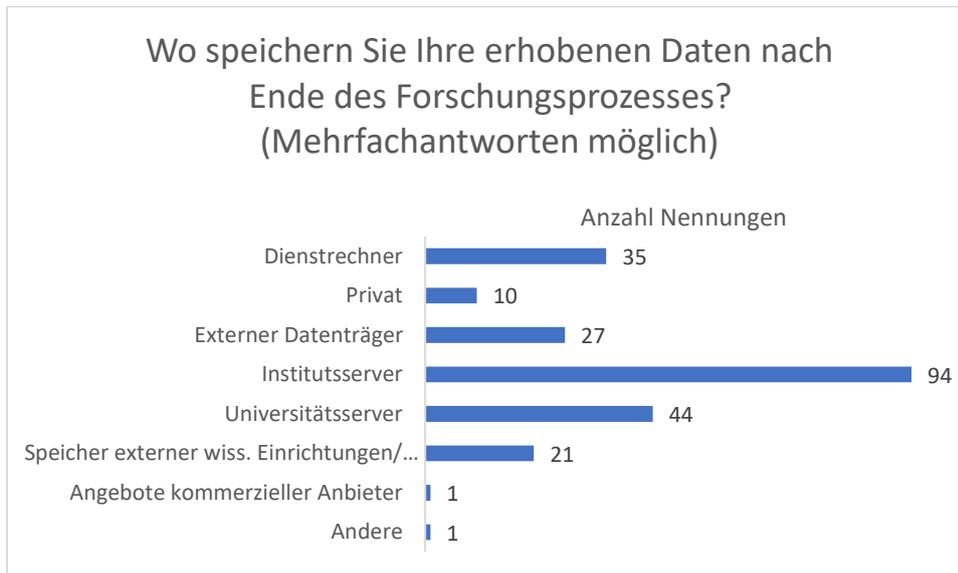


Abbildung 22 „Wo speichern Sie Ihre erhobenen Daten nach Ende des Forschungsprozesses?“ Mehrfachnennungen möglich (n=129, Anzahl Nennungen 233)

Auf die Frage nach dem Vorhandensein passender Repositorien antwortet die Hälfte der Befragten mit *weiß nicht*.



Abbildung 23 „Gibt es passende Repositorien für Ihre Forschungsdaten?“ (n=138)

## Zusammenfassung

Im Hinblick auf das Forschungsdatenmanagement lässt sich konstatieren, dass das Gros der Befragten personenbezogene Daten selber erhebt und/oder Daten, die von anderen erhoben wurden, verwaltet. Die Erstellung von Datenmanagementplänen hat sich hierbei noch nicht richtig durchgesetzt. Generell kann von einer Standardisierung im Bereich Forschungsdatenmanagement an Institutionen bzw. Projekten der Befragten nicht gesprochen werden. Was die jeweiligen Standards für die Sicherung, Erfassung, Formatierung und Beschreibung von Forschungsdaten betrifft, herrscht zudem bei der Hälfte Unwissen über deren Vorhandensein. Die für das Forschungsdatenmanagement genutzten Tools sind sehr heterogen, trotzdem ist Interoperabilität noch kein Thema. Die Speicherung der Daten erfolgt sowohl während als auch nach dem Forschungsprozess auf lokalen Servern. Über das Vorhandensein von passenden Repositorien besteht noch eine Wissenslücke.

Wenn man von der naheliegenden These ausgeht, dass diejenigen, die sich an der Befragung beteiligt haben, tendenziell eher Forschende sind, die bereits eine höhere Affinität zum Thema Forschungsdatenmanagement haben, werden die tatsächlichen Desiderate im Bereich der medizinischen Forschungsdaten eher noch gravierender sein, als sich aus dieser Befragung ableiten lässt.

## Forschungsdatenbereitstellung

Im Fragenblock *Forschungsdatenbereitstellung* werden die Bereitschaft und Motivation zum Teilen von Daten, aber auch dabei entstehenden Herausforderungen behandelt. Es werden auch hier zu Beginn Filter verwendet, um diejenigen weiterzuleiten, die entweder noch nie mit jemandem Daten geteilt haben oder dies nicht wollen. Für die befragte Grundgesamtheit traf dies nur für 18 Personen zu. Sie übersprangen den Teil mit den Fragen zur Forschungsdatenbereitstellung und wurden zu einem Fragenblock weitergeleitet, der ihre Gründe bzw. Motivation hierfür erhebt. Da hier aufgrund der geringen Fallzahl derjenigen, die diesen Fragenkomplex ausfüllten (n=11), keine aussagekräftigen Ergebnisse vorliegen, wird dieser Teil bewusst ausgelassen.

Die Befragten teilen primär aufbereitete Daten mit Dritten (87 Nennungen bei n=134), aber auch Erhebungsinstrumente (69 Nennungen) und Studiendokumente (65 Nennungen).



Abbildung 24 „Was haben Sie schon einmal mit Dritten geteilt?“ Mehrfachnennungen möglich (n=134, Anzahl Nennungen 480)

Geteilt wird unter den Befragten am ehesten mit Forschenden innerhalb der eigenen Institution (97 Nennungen bei n=134), aber auch mit Forschenden, die persönlich bekannt sind (77 Nennungen).



Abbildung 25 „Wem haben Sie selbst schon einmal Ihre Daten zur Verfügung gestellt?“ Mehrfachnennungen möglich (n=134, Anzahl Nennungen 293)

Dies schlägt sich auch in der Art und Weise wieder, in der die Daten u.Ä. geteilt bzw. zur Verfügung gestellt wurden: Dies erfolgt zumeist über persönliche Kommunikationswege (69 Nennungen bei n=134), Downloadserver (46 Nennungen) und/oder als vor Ort gestatteter Zugang (37 Nennungen).

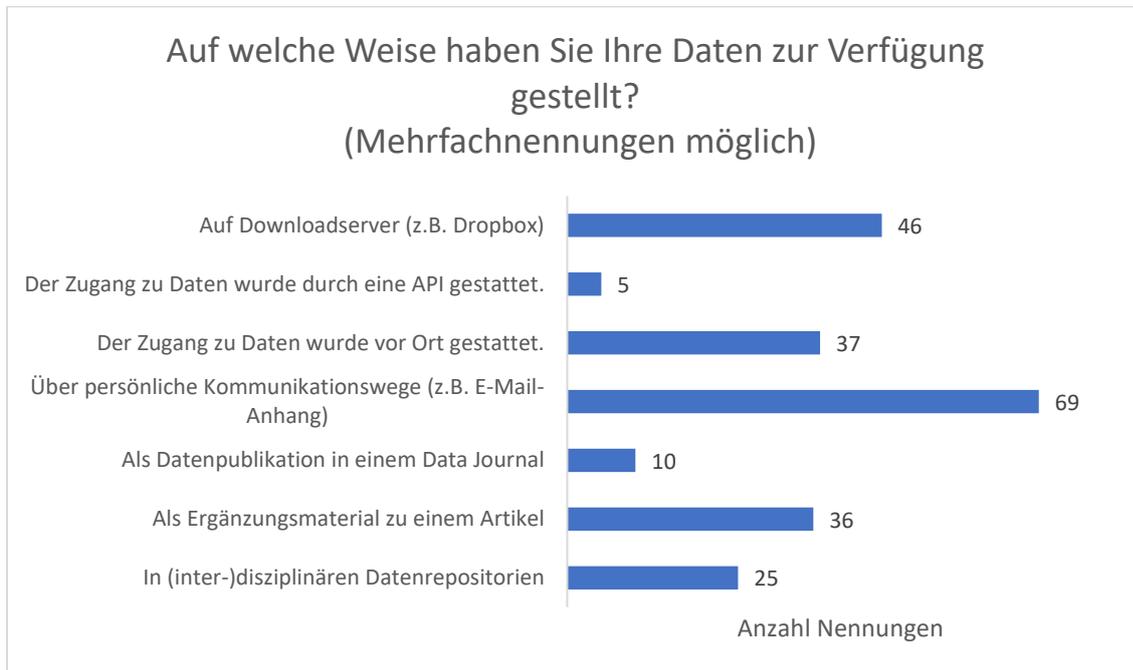


Abbildung 26 „Auf welche Weise haben Sie Ihre Daten zur Verfügung gestellt?“ Mehrfachnennungen möglich (n=134, Anzahl Nennungen 228)

Im nächsten Schritt wurde die Motivation für das Teilen von Daten abgefragt. Die Teilnehmenden sollten ihre Zustimmung zu verschiedenen Aussagen geben. Daten werden Dritten zur Verfügung gestellt, um Kooperationen mit anderen zu begünstigen (83 % Zustimmung), um die Validierung und Verifizierung von Forschungsergebnissen zu unterstützen (75 % Zustimmung) sowie um das eigene Forschungsgebiet voranzubringen und weil andere wichtige Forschende ihre Daten ebenfalls teilen (jeweils 62 % Zustimmung). Interessant in diesem Kontext ist auch die negative Bewertung der Aussage, das Teilen von Daten erhöhe die Chancen auf dem Arbeitsmarkt – ihr stimmen 61 % der Befragten nicht zu.

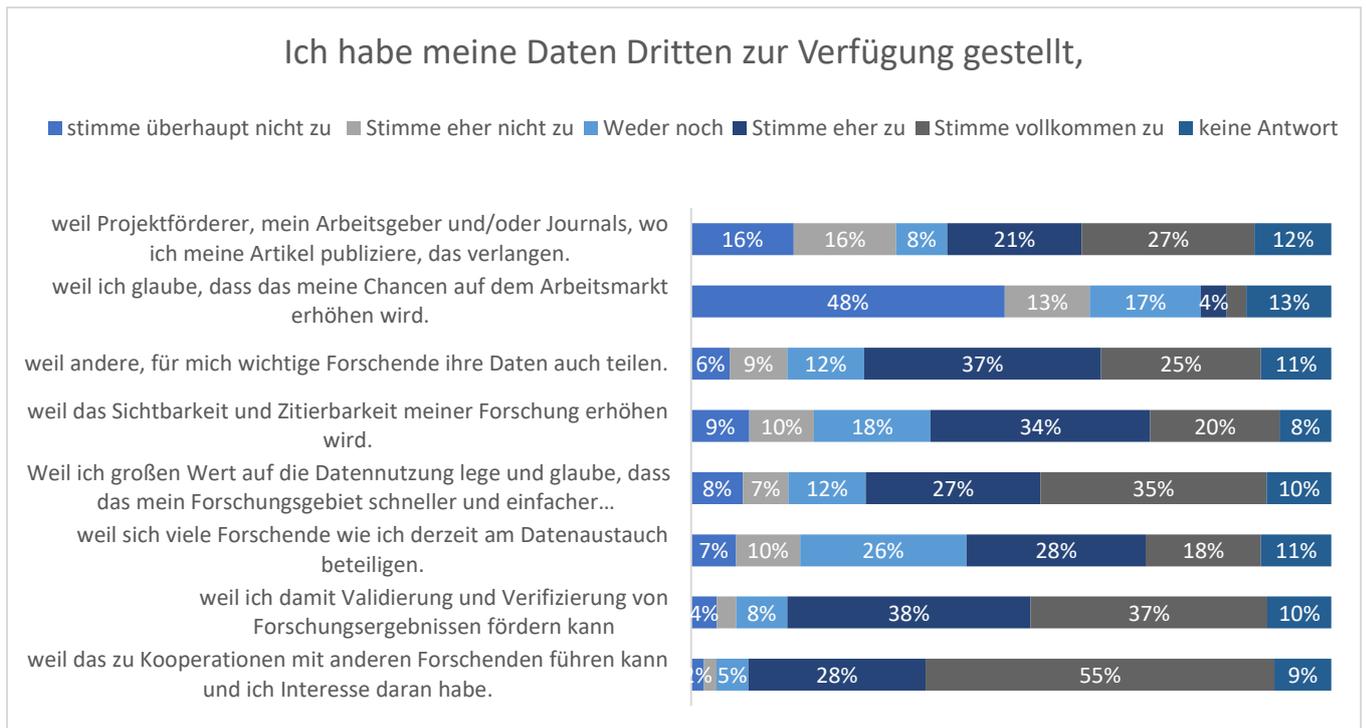


Abbildung 27 „Ich habe meine Daten Dritten zur Verfügung gestellt?“ (n=101)

Bei der Frage danach, wem die Teilnehmenden ihre Daten potentiell noch zur Verfügung stellen würden, ergibt sich ein ähnliches Bild wie bei jener nach dem Personenkreis, dem bereits Daten zur Verfügung gestellt wurden. In erster Linie würden die Befragten Forschenden innerhalb der eigenen Einrichtung Daten zur Verfügung stellen (115 Nennungen bei n=125), aber auch Forschenden außerhalb der Einrichtung, zu denen eine persönliche Verbindung besteht (97 Nennungen), oder solchen, die an einem ähnlichen Thema arbeiten (91 Nennungen).

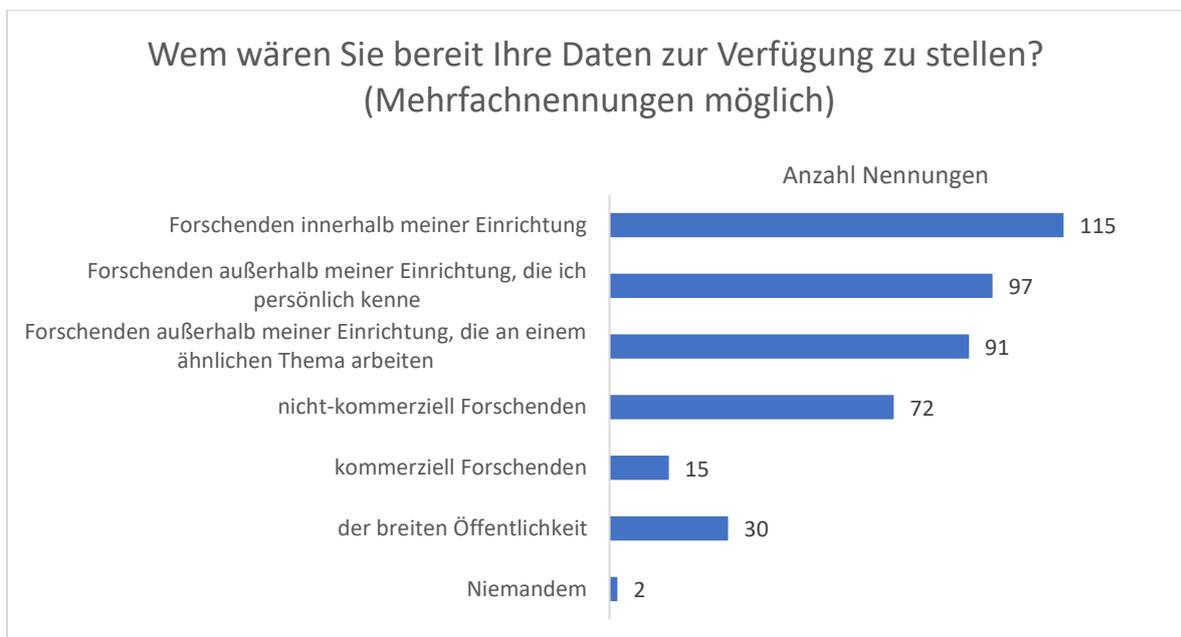


Abbildung 28 „Wem wären Sie bereit Ihre Daten zur Verfügung zu stellen?“ Mehrfachnennungen möglich (n=125, Anzahl Nennungen 442)

Dabei ist der Wille, mit anderen Forschenden auch außerhalb der eigenen Institution zu teilen, durchaus vorhanden. Es besteht eine hohe Bereitschaft, Erhebungsinstrumente (z.B. Fragebogenmuster; 80 Nennungen bei n=125) und/oder Metadaten (78 Nennungen) mit Dritten zu teilen.

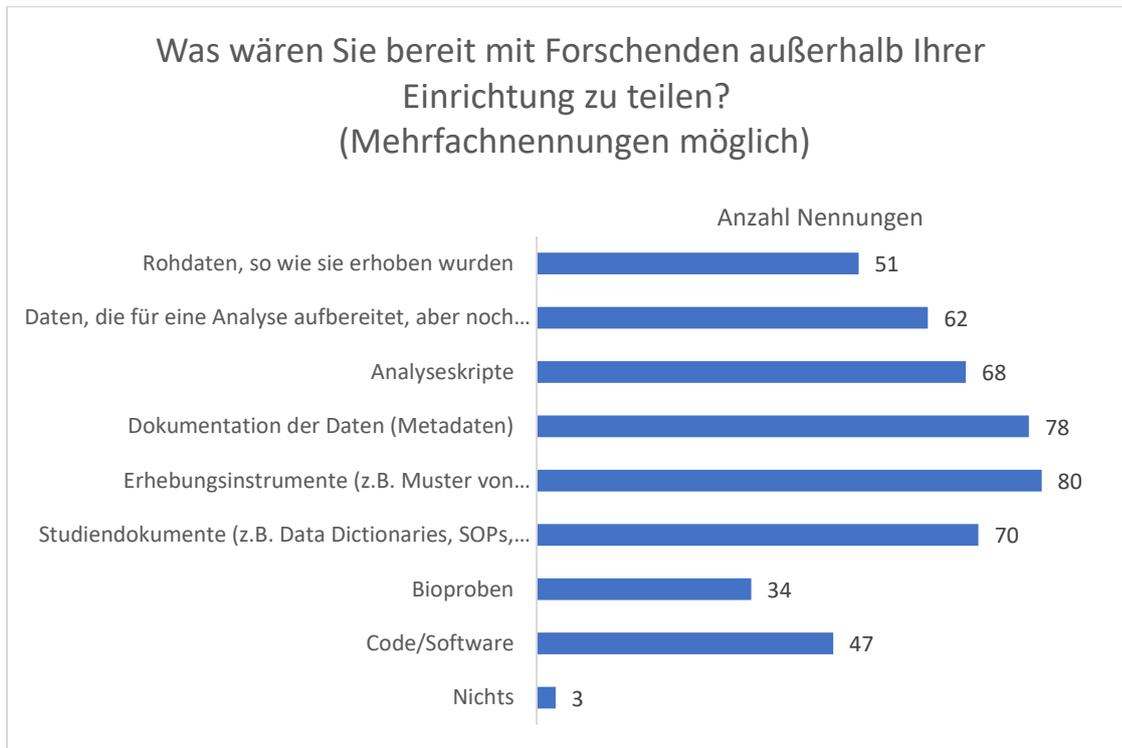


Abbildung 29 „Was wären Sie bereit mit Forschenden außerhalb Ihrer Einrichtung zu teilen?“ Mehrfachnennungen möglich (n=125, Anzahl Nennungen 493)

Die Befragten sind im Gros nur dann bereit, ihre Daten Dritten zur Verfügung zu stellen, wenn die Datenweitergabe auf Basis einer konkreten Nutzungsabsprache erfolgt (97 Nennungen bei n=125) oder wenn die Forschungsdaten formal zitiert werden (84 Nennungen).

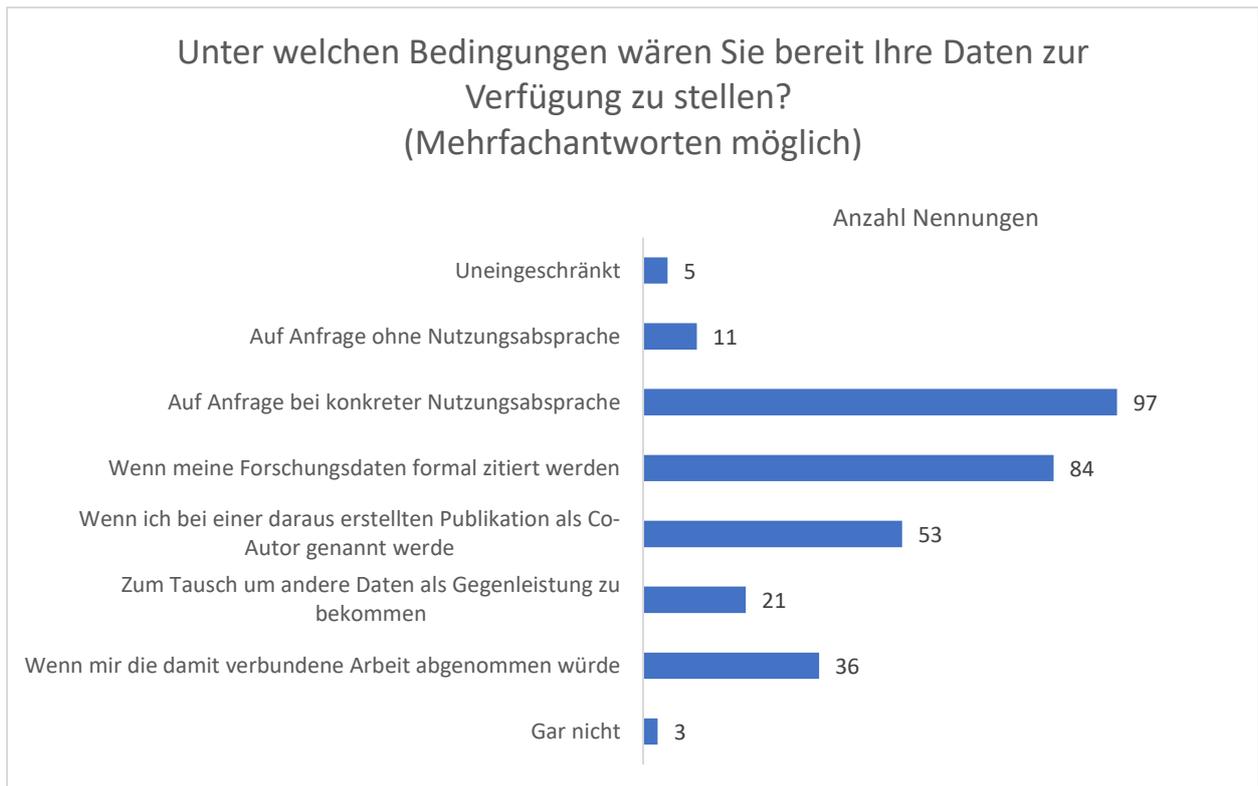


Abbildung 30 „Unter welchen Bedingungen wären Sie bereit Ihre Daten zur Verfügung zu stellen?“ Mehrfachnennungen möglich (n=125, Anzahl Nennungen 310)

Es besteht zudem unter den Befragten die Bereitschaft, Studiendaten für Studienteilnehmende primär in aggregierter Form als Übersicht (84 Nennungen bei n=112) digital zur Verfügung zu stellen.

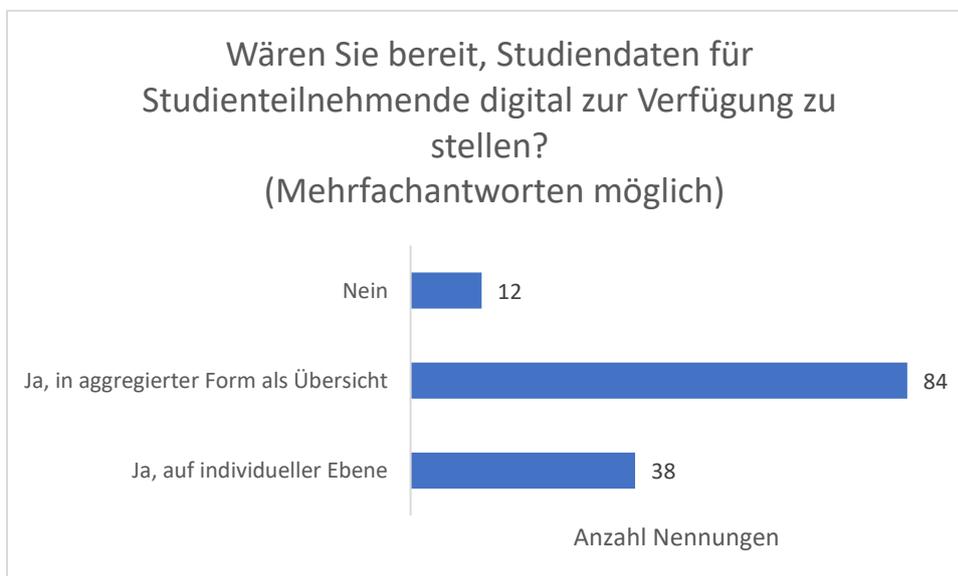


Abbildung 31 „Wären Sie bereit, Studiendaten für Studienteilnehmende digital zur Verfügung zu stellen?“ Mehrfachnennungen möglich (n=112, Anzahl Nennungen 134)

Dabei schätzen 43 % der Befragten das Interesse von Studienteilnehmenden, sich über Daten der Studie, an der sie teilnehmen, zu informieren und ggf. auch selber Vorschläge für neue Studienfragen und -auswertungen zu machen, auf einer Skala von 1 (nicht vorhanden) bis 5 (sehr hoch) zwischen 4 und 5 ein.

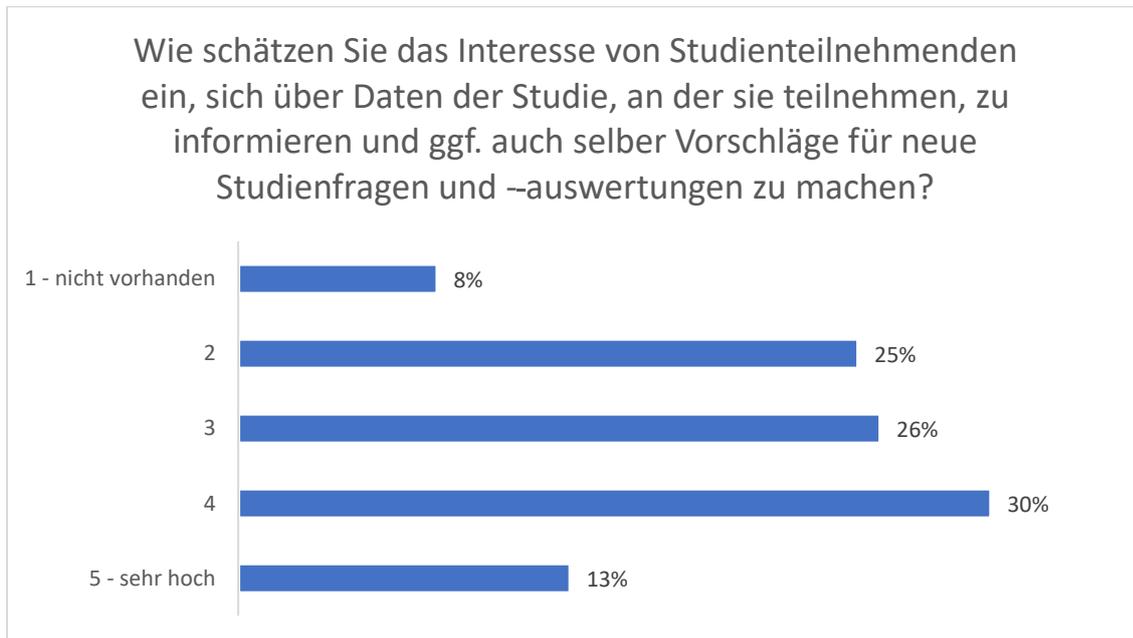


Abbildung 32 „Wie schätzen Sie das Interesse von Studienteilnehmenden ein, sich über Daten der Studie, an der sie teilnehmen, zu informieren und ggf. auch selber Vorschläge für neue Studienfragen und -auswertungen zu machen?“ (n=113)

Bei der Bereitstellung der Daten geben 38 % der Teilnehmenden an, immer auf deren Maschinenverarbeitbarkeit zu achten; 34 % tun dies meistens.

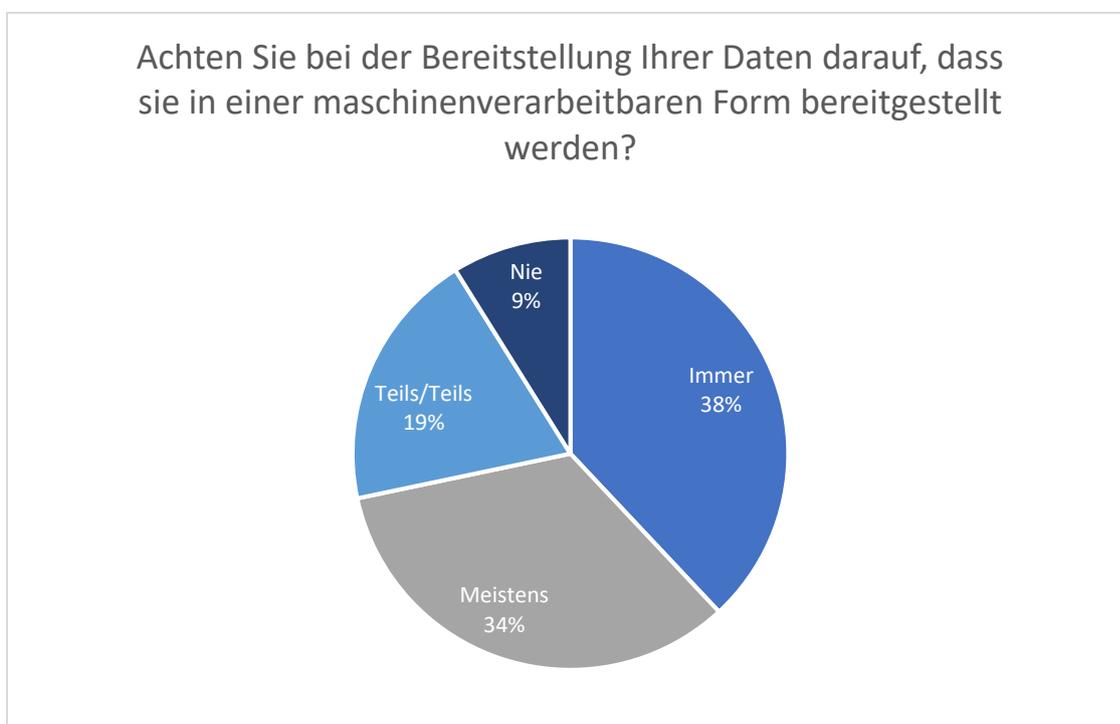


Abbildung 33 „Achten Sie bei der Bereitstellung Ihrer Daten darauf, dass sie in einer maschinenverarbeitbaren Form bereitgestellt werden?“ (n=111)

Im abschließenden Teil dieses Fragenkomplexes wurden die Teilnehmenden gebeten anzugeben, inwieweit eine Reihe von Aussagen auf sie zutreffen. So sind 44 % der Befragten der Ansicht, über ausreichend Kenntnisse und Kompetenzen zu verfügen, um die zu teilenden Daten zu organisieren, allerdings haben 45 % juristische Bedenken, die Daten zur Nachnutzung zur Verfügung zu stellen. 39 % der Befragten gibt an, dass ihre Institutionen über geeignete Informationssysteme, gut funktionierende technologische Infrastruktur und ausreichende Netzwerkkapazitäten verfügen, um effiziente Datenbereitstellung zu ermöglichen. Für 54 % trifft es mindestens zu, dass auch passende Repositorien zur Verfügung gestellt werden. Die Erhöhung der Zitierraten (27 %) sowie die Erhöhung der Reputation (28 %) sind nur für einen kleineren Teil der Befragten ein Motivator dafür, Daten zur Verfügung zu stellen. Lediglich 12 % stimmen der Aussage zu, dass ihre Daten missinterpretiert und/oder missbraucht werden könnten.

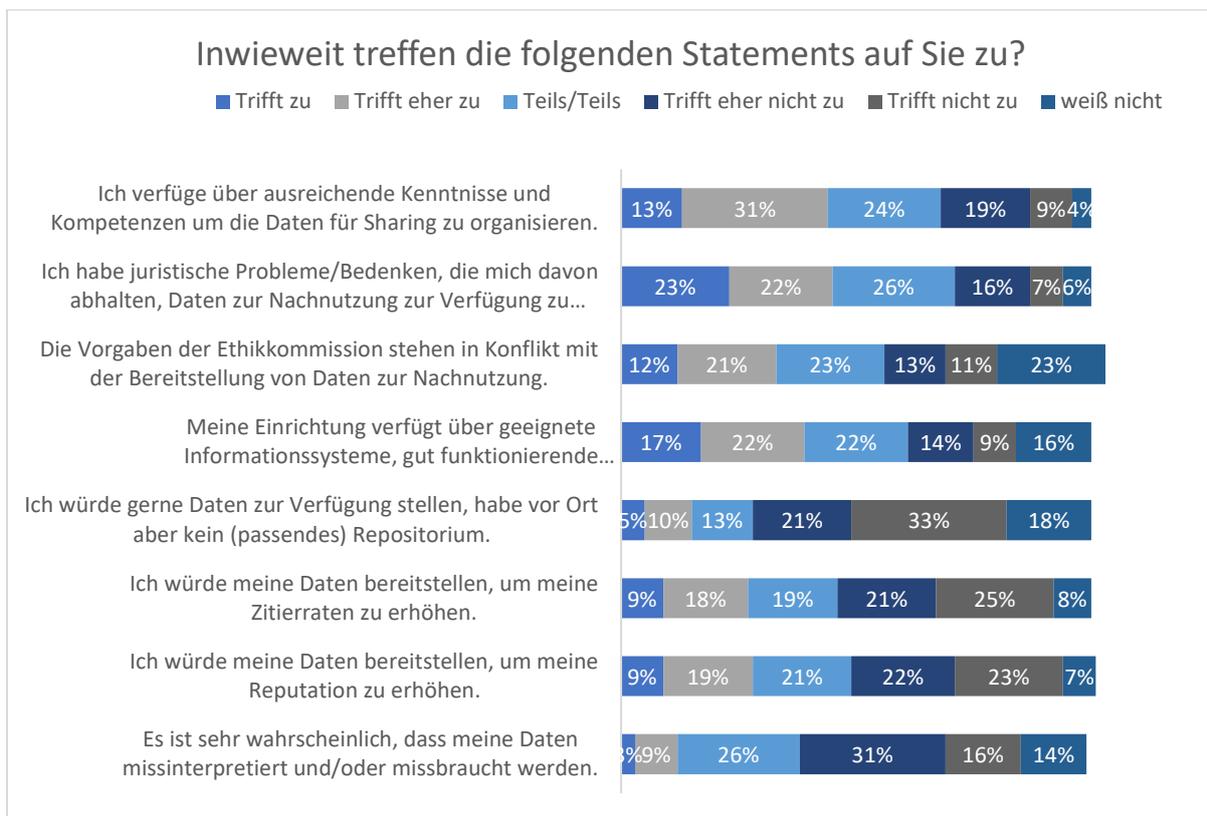


Abbildung 34 „Inwieweit treffen die folgenden Statements auf Sie zu?“ (n=117)

In der abschließenden offenen Frage wurde nochmals explizit danach gefragt, was die Teilnehmenden sonst noch für die maschinenverarbeitbare Bereitstellung von Daten benötigen. Hier äußern diejenigen, die das Feld ausgefüllt haben (n=19), den Wunsch nach rechtlichen Guidelines (10) und Schulungen bzw. mehr Informationen (5).

### Zusammenfassung

In der befragten Grundgesamtheit herrscht (vermutlich auch durch Mechanismen der Selbstselektion) eine große Bereitschaft, Daten mit Dritten zu teilen. Die meisten Befragten haben bereits Daten und Erhebungsinstrumente – meistens mit persönlich bekannten Forschenden bzw. Personen – geteilt. Dies geschah primär über persönliche Kanäle und/oder persönlich gewährten Zugang. Der Befund zeigt, dass in diesem Bereich noch bislang ungenutzte Potentiale schlummern und die Forschenden hier mehr Informationen und sichere Plattformen für den Austausch benötigen.

Die Motivation zum Teilen entsteht zum einen aus dem Wunsch, Kooperationen zwischen Forschenden, aber auch die Validierung und Verifizierung von Forschungsergebnissen, zu fördern und das eigene Forschungsgebiet voranzubringen. Zum anderen ist die dadurch entstehende

Reziprozität ein Motiv, weil andere ihre Daten im Gegenzug ebenfalls teilen. Was bei der Motivation, Daten Dritten zur Verfügung zu stellen, eine geringere Rolle spielt, ist die Aussicht auf Reputationsgewinn oder bessere Chancen auf dem Arbeitsmarkt. Allerdings bestehen bei mehr als der Hälfte der Befragten juristische Bedenken hinsichtlich der Bereitstellung von Daten. Hier kann NFDI4Health einen wichtigen Beitrag leisten und die bereits vorhandene Motivation der Community, Daten zu teilen, mit Hilfe von Schulungen und Guidelines noch erhöhen.

Prinzipiell würden noch mehr Befragte Daten mit Personen auch außerhalb der eigenen Forschungseinrichtung teilen, wenn im Vorfeld konkrete Nutzungsabsprachen getroffen würden und/oder eine formale Zitation der Daten erfolgte. Auch besteht die Bereitschaft, Studienteilnehmenden Daten zur Verfügung zu stellen, allerdings wünschen sich die Teilnehmenden hierzu noch rechtliche Grundlagen bzw. Guidelines und Schulungen. Dies zeigt, dass NFDI4Health mit den Zielen, Schulungen zum Thema Forschungsdaten anzubieten, den Wünschen der Community entspricht und dass der Bedarf gegeben ist. Ähnlich verhält es sich mit dem Wunsch nach rechtlichen Grundlagen und Musterformularen für das Teilen von Daten, der ebenfalls durch das Projekt adressiert wird.

### Forschungsdatennutzung

Der nächste Fragenkomplex behandelt den Aspekt der (Nach-)Nutzung von Forschungsdaten, die von Dritten erhoben wurden. In der befragten Grundgesamtheit nutzen weniger als die Hälfte der Befragten Daten nach, weshalb die Fallzahlen für die Teilfragen aus diesem Abschnitt im zweistelligen Bereich liegen, aber trotzdem wichtige Tendenzen aufzeigen.

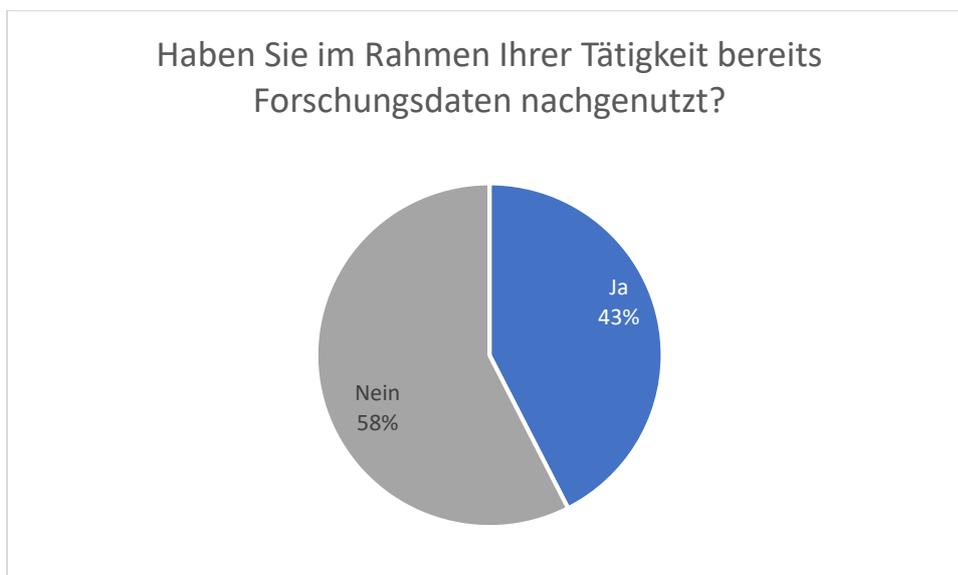


Abbildung 35 „Haben Sie im Rahmen Ihrer Tätigkeit bereits Forschungsdaten nachgenutzt?“ (n=120)

Diejenigen, die Daten nachnutzen, suchen nach diesen in erster Linie in wissenschaftlichen Zeitschriften (30 Nennungen bei n=50) oder auf Webseiten relevanter Forschungsprojekte (29 Nennungen).

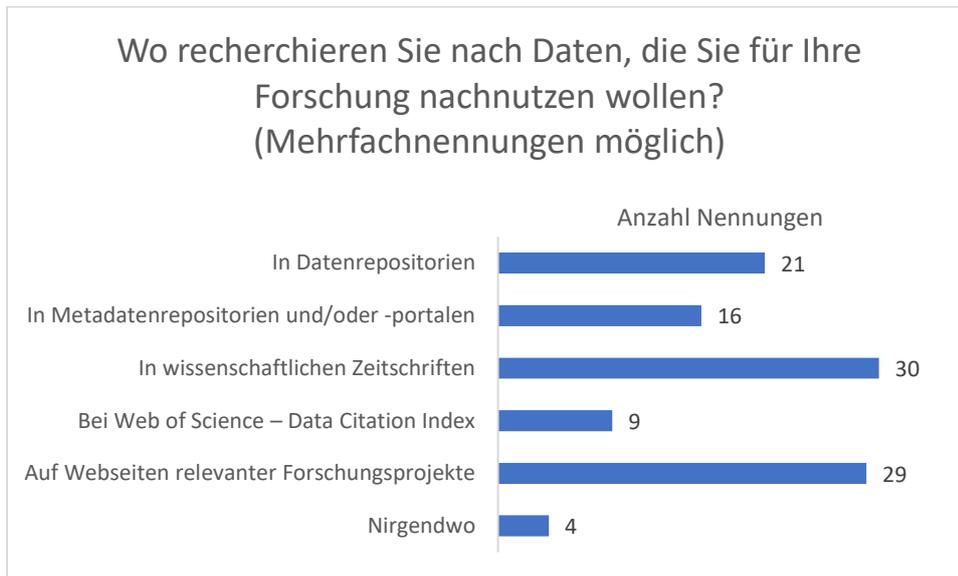


Abbildung 36 „Wo recherchieren Sie nach Daten, die Sie für Ihre Forschung nachnutzen wollen?“ Mehrfachnennungen möglich (n=50, Anzahl Nennungen 109)

Bei der Zufriedenheit hinsichtlich der bisherigen Möglichkeiten, Daten zu finden, verortet sich knapp die Hälfte bei der Antwortoption „neutral“.

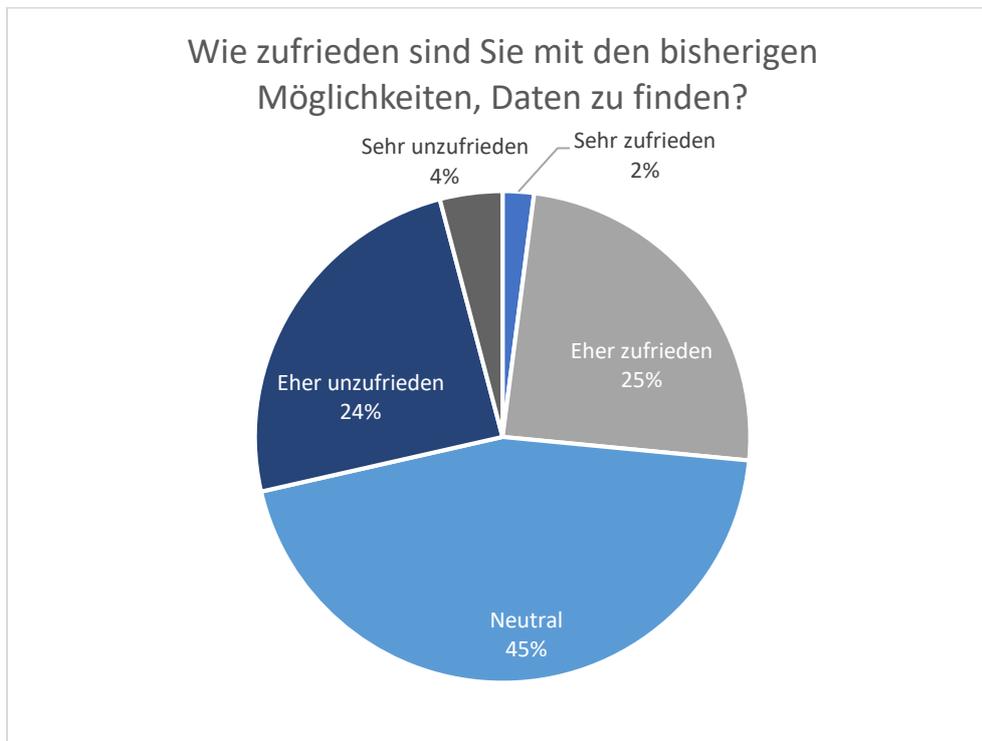


Abbildung 37 „Wie zufrieden sind Sie mit den bisherigen Möglichkeiten, Daten zu finden?“ (n=48)

Die Daten werden primär für die Formulierung eigener Fragestellungen genutzt (36 Nennungen bei n=48), um einen Überblick zu gewinnen (29 Nennungen) oder zur Reproduktion und Validierung bzw. Evaluation von Forschungsdaten (27 Nennungen).

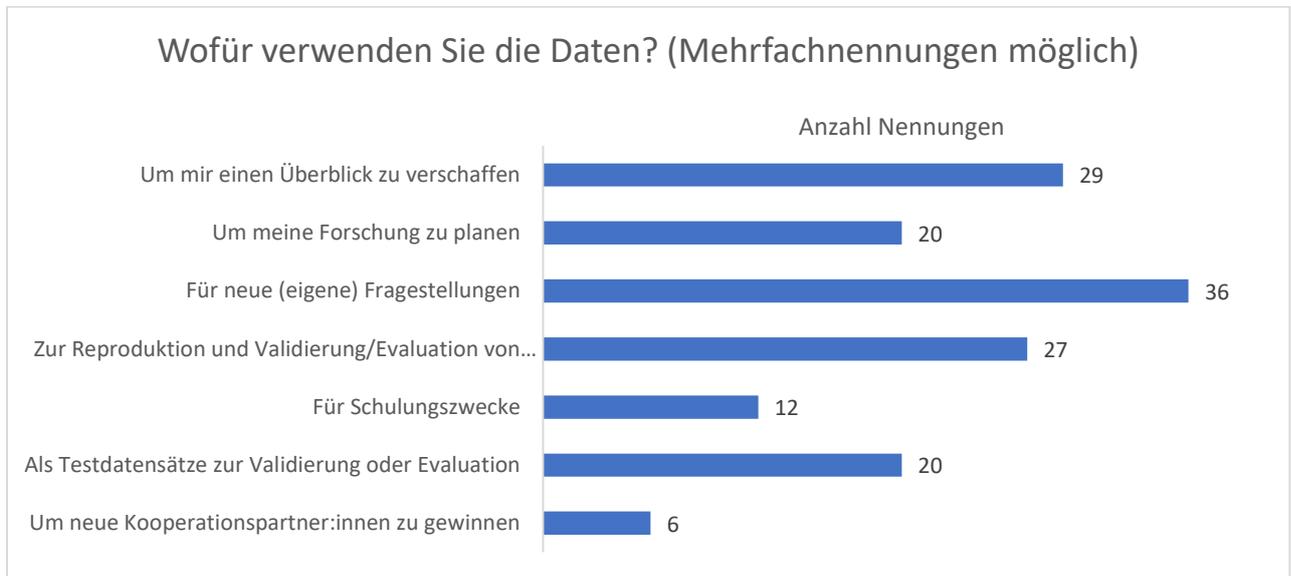


Abbildung 38 „Wofür verwenden Sie die Daten?“ Mehrfachnennungen möglich (n=48, Anzahl Nennungen 150)

Das Feld der von Dritten benötigten Daten ist sehr heterogen, allerdings stechen hier die strukturierten numerischen Daten klar heraus (46 Nennungen bei n=160).

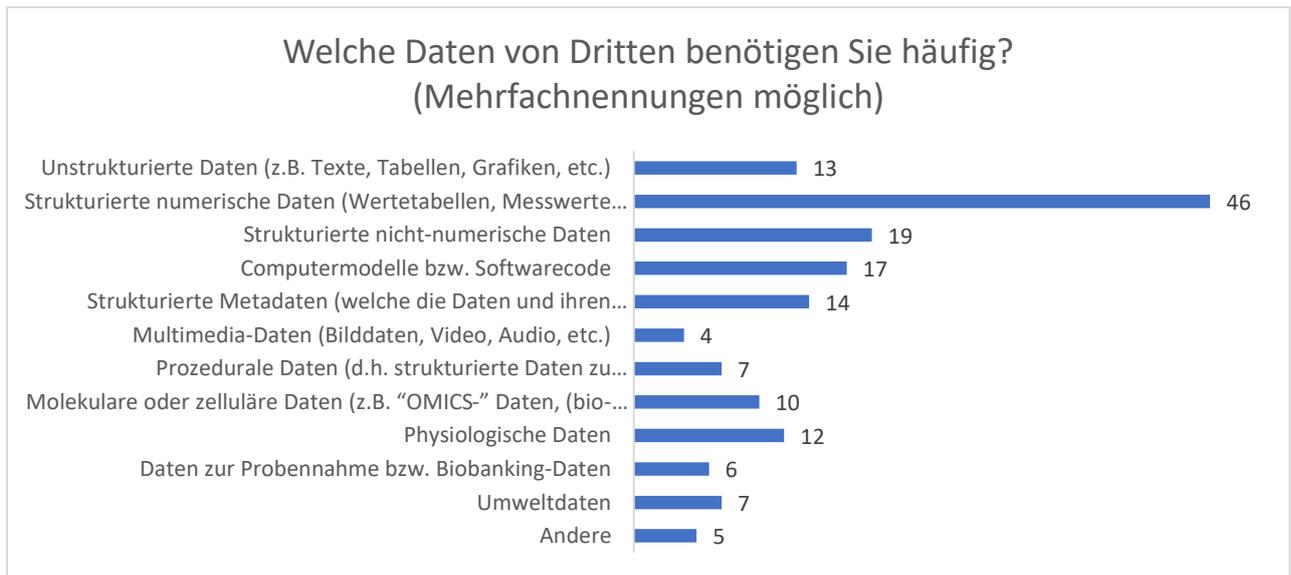


Abbildung 39 „Welche Daten von Dritten benötigen Sie häufig?“ Mehrfachnennungen möglich (n=48, Anzahl Nennungen 160)

Hinsichtlich der Herausforderungen und Schwierigkeiten bei der Nachnutzung fremder Forschungsdaten werden insbesondere die ungenügende Beschreibung der Daten (62 %), ein kompliziertes Prozedere bei der Beschaffung (58 %) und unklare Situationen hinsichtlich Lizenzrechten und/oder Datenschutz (38 %) als gravierend empfunden.

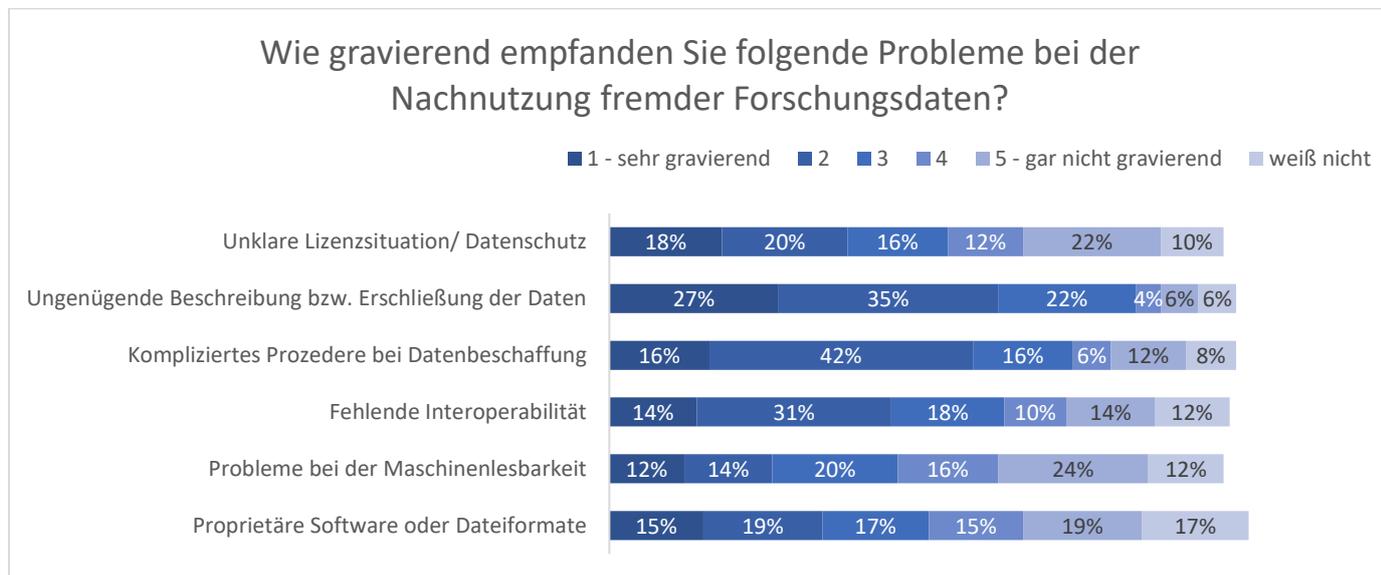


Abbildung 40 „Wie zufrieden sind Sie mit den bisherigen Möglichkeiten, Daten zu finden?“ (n=49)

Die nachfolgenden Fragen wurden allen am Fragebogen Teilnehmenden gestellt, weshalb hier die n-Werte auch wieder in den dreistelligen Bereich gehen.

Generell ist das Interesse an der Nachnutzung aller zur Auswahl stehender (Forschungs-)Materialien hoch bis sehr hoch. Besonderes Interesse besteht an erhobenen Daten (72 %) und Erhebungsinstrumenten (70 %).

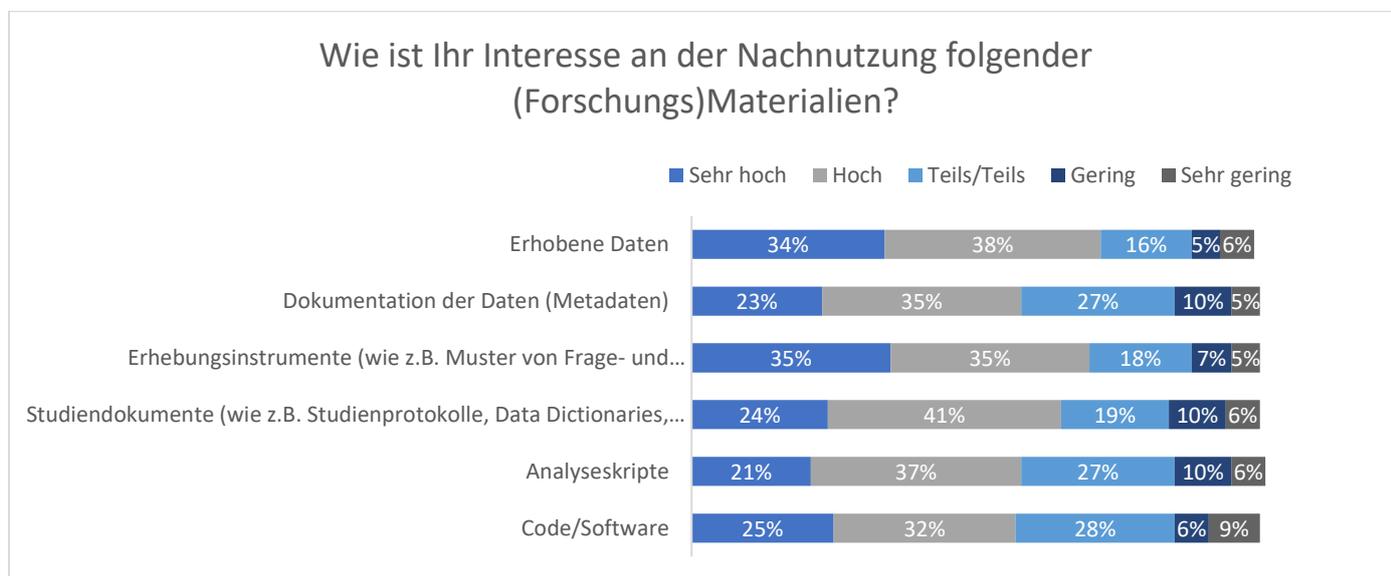


Abbildung 41 „Wie ist Ihr Interesse an der Nachnutzung folgender (Forschungs)Materialien?“ (n=109)

In der abschließenden offenen Frage wurde nochmals explizit danach gefragt, was die Teilnehmenden weiterhin noch für die maschinenverarbeitbare Bereitstellung von Daten benötigen. Hier äußern jene, die das Feld ausgefüllt haben (n=19), den Wunsch nach rechtlichen Guidelines (10) und Schulungen bzw. mehr Informationen (5).

## Zusammenfassung

Die befragte Grundgesamtheit teilt häufiger eigene Forschungsdaten als sie Forschungsdaten von Dritten nachnutzt. Daten von Dritten – primär strukturierte numerische Daten – werden dazu genutzt, einen Überblick über ein Thema zu erhalten oder um die eigene Fragestellung zu schärfen. Dabei begegnen den Forschenden Hürden in Form von ungenügender Beschreibung der Daten, kompliziertem Prozedere bei der Beschaffung und/oder unklaren Situationen hinsichtlich Lizenzierung bzw. Datenschutz. Dabei besteht bei der Grundgesamtheit der Befragten durchaus großes Interesse an der Nachnutzung von (Forschungs-)Materialien.

Auch die Ergebnisse dieses Fragenkomplexes zeigen, dass NFDI4Health mit seinen Zielen, das Teilen von Daten zu erleichtern sowie ihre Nachnutzung einfacher zu gestalten, jenen entgegenkommt, die bereits Daten nachnutzen. Wenn das Prozedere vereinfacht und FAIR-Standards implementiert werden, sollte die Erschließung und Beschreibung der Daten zukünftig kein Problem mehr darstellen. Zusätzlich sei noch erwähnt, dass dadurch auch die Motivation, Daten nachzunutzen und Synergien zu finden, verstärkt werden könnte.

## Schulungsbedarf

Der in den vorhergehenden Fragenblöcken von den Befragten implizit geäußerte Bedarf an Schulungen wurde im folgenden Fragenkomplex behandelt. Auch hier wurde zu Beginn eine Filterfrage gesetzt, die jene, die an Schulungen interessiert waren, von Nichtinteressierten trennte. Etwas weniger als zwei Drittel der Befragten gaben an, noch zusätzliche Schulungen bzw. Informationen rund um das Thema Forschungsdaten zu benötigen.

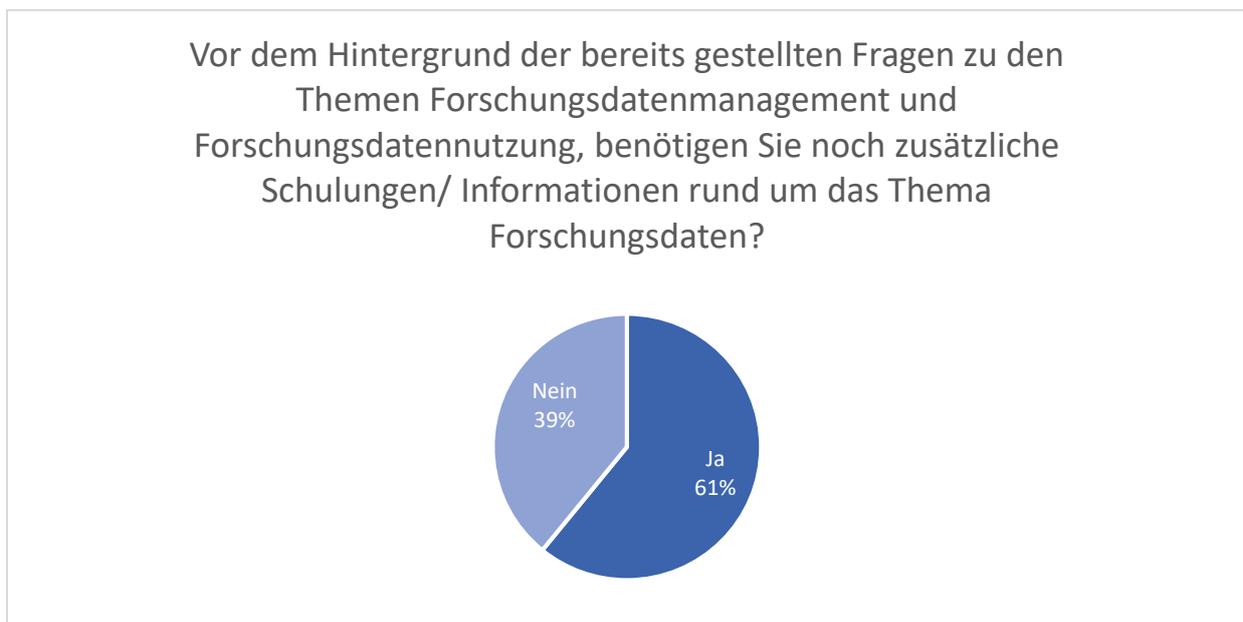


Abbildung 42 „Vor dem Hintergrund der bereits gestellten Fragen zu den Themen Forschungsdatenmanagement und Forschungsdatennutzung, benötigen Sie noch zusätzliche Schulungen/ Informationen rund um das Thema Forschungsdaten?“ (n=115)

Diejenigen, die ihr Interesse bekundeten, wurden im nächsten Schritt dazu aufgefordert, die für sie relevanten Themen und das gewünschte Niveau der Schulung zu benennen. Bei den Themen *Daten nachnutzen* (41 %), *Daten archivieren* (42 %) und *Datenschutz* (37 %) besteht eher Interesse an Schulungen bzw. Informationen auf Einsteigerniveau und etwas darüber, während die Befragten bei den Themen *Daten aufbereiten und analysieren* (60 %), *Daten erheben* (54 %) und *Forschungsvorhaben planen* (52 %) eher auf Expertenniveau informiert bzw. weitergebildet werden möchten.

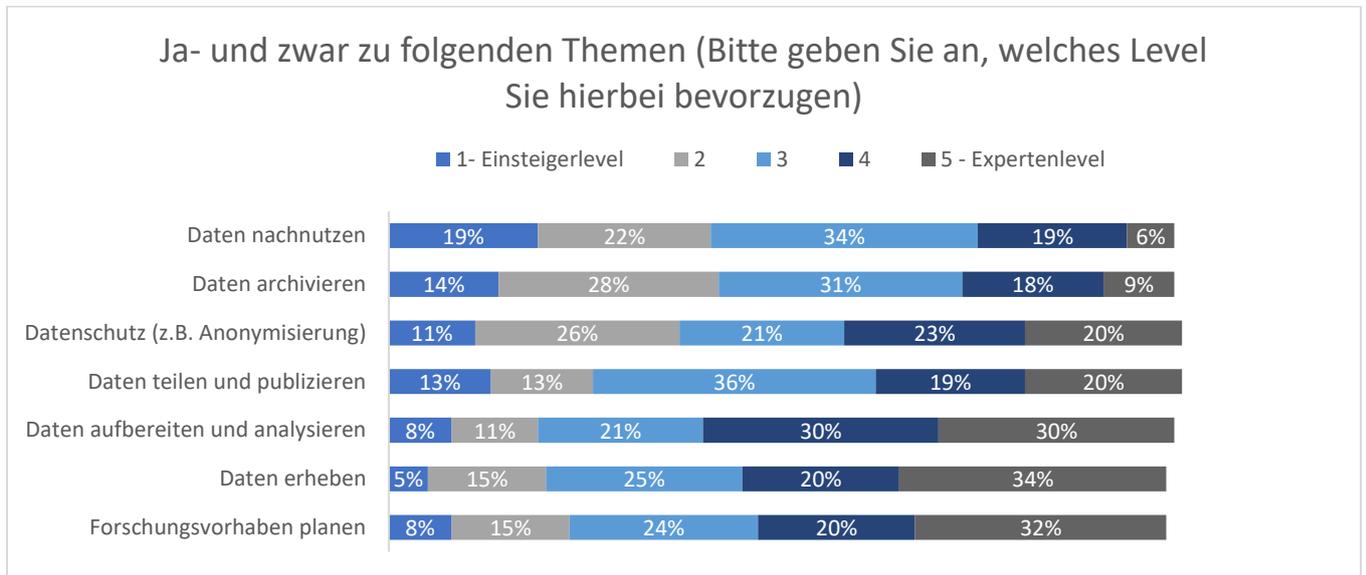


Abbildung 43 „Ja- und zwar zu folgenden Themen (Bitte geben Sie an, welches Level Sie hierbei bevorzugen)“ (n=109)

Ihren Informationsbedarf möchten die Teilnehmenden primär durch Seminare decken (56 Nennungen bei n=64), aber auch Online-Infos (40 Nennungen) und Selbstlernutorials (39 Nennungen) sind gefragt.

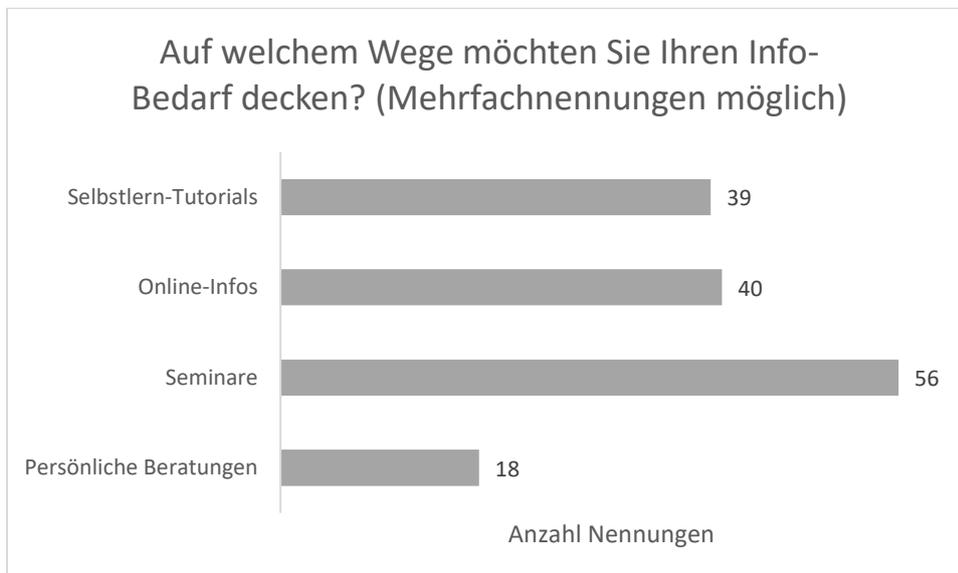


Abbildung 44 „Auf welchem Wege möchten Sie Ihren Info-Bedarf decken?“ Mehrfachnennungen möglich (n=64, 153 Nennungen)

Anschließend wurden erneut alle Teilnehmenden nach dem Vorhandensein einer zentralen Anlaufstelle für Forschungsdatenmanagement an ihrer Institution befragt. 60 % geben an, dass es an ihrer Institution keine solche Stelle gibt.

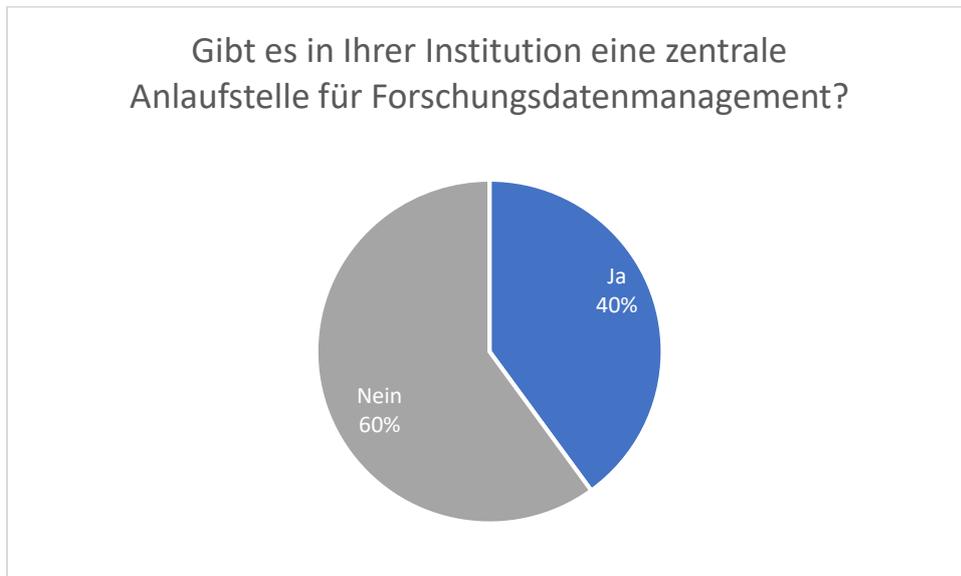


Abbildung 45 „Gibt es in Ihrer Institution eine zentrale Anlaufstelle für Forschungsdatenmanagement?“ (n=115)

Von denjenigen, die diese Frage bejahten, hatten 76 % bereits Kontakt zu der entsprechenden zentralen Anlaufstelle in ihrer Institution.

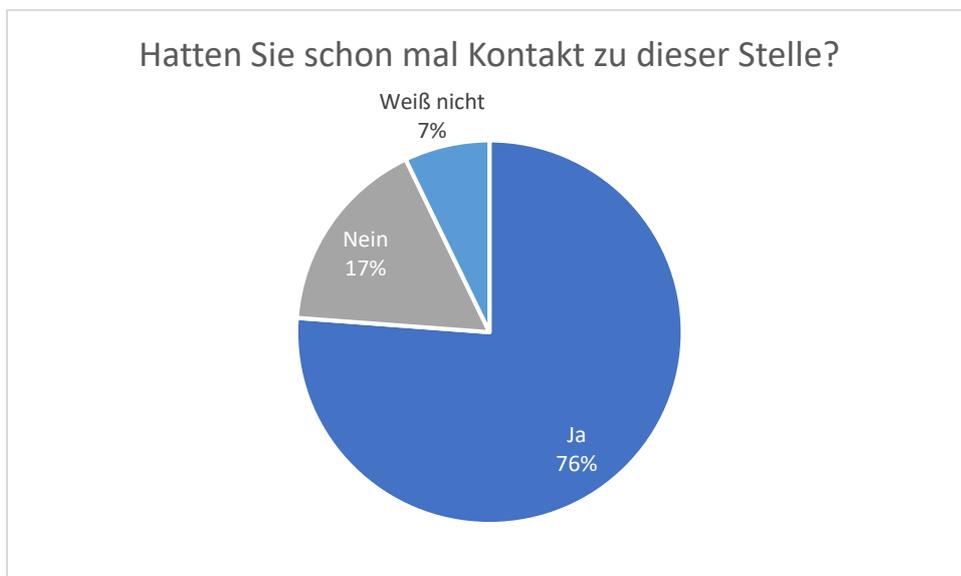


Abbildung 46 „Hatten Sie schon mal Kontakt zu dieser Stelle?“ (n=42)

### Zusammenfassung

Bei den Fragen nach dem Bedarf an Schulungen zeigt sich, dass in diesem Bereich eine Nachfrage existiert. Abhängig vom Thema wünschen sich die Befragten Informationen entweder eher auf Einsteiger- oder auf Expertenniveau; bevorzugt in Seminarform, aber auch als Online-Infos und/oder Selbstlerntutorials.

Der Großteil der Befragten gibt an, keine zentrale Anlaufstelle für Forschungsdatenmanagement in der jeweiligen Institution zu haben. Wenn es eine gibt, besteht in den meisten Fällen bereits Kontakt zu ihr.

Die Ziele von NFDI4Health, Guidelines und Schulungen zum Thema Forschungsdatenmanagement zu implementieren, adressieren einen klaren Wunsch der Community, die sich insbesondere hinsichtlich der juristischen Rahmenbedingungen oftmals noch unsicher ist. Auch kann NFDI4Health mit seiner Expertise eine Lücke füllen und als zentrale Anlaufstelle für Fragen rund um Forschungsdatenmanagement fungieren.

## NFDI4Health

Im folgenden Fragenteil sollten die Bekanntheit, aber auch die Inhalte von NFDI4Health abgefragt werden.

Die erste Frage betraf die Bewertung der im Antrag hervorgehobenen Ziele von NFDI4Health. Jeder Punkt erhielt eine hohe Anzahl an Nennungen. Die Befragten hatten außerdem die Möglichkeit, weitere Anforderungen zu formulieren, wovon aber nur in sieben Fällen Gebrauch gemacht wurde.

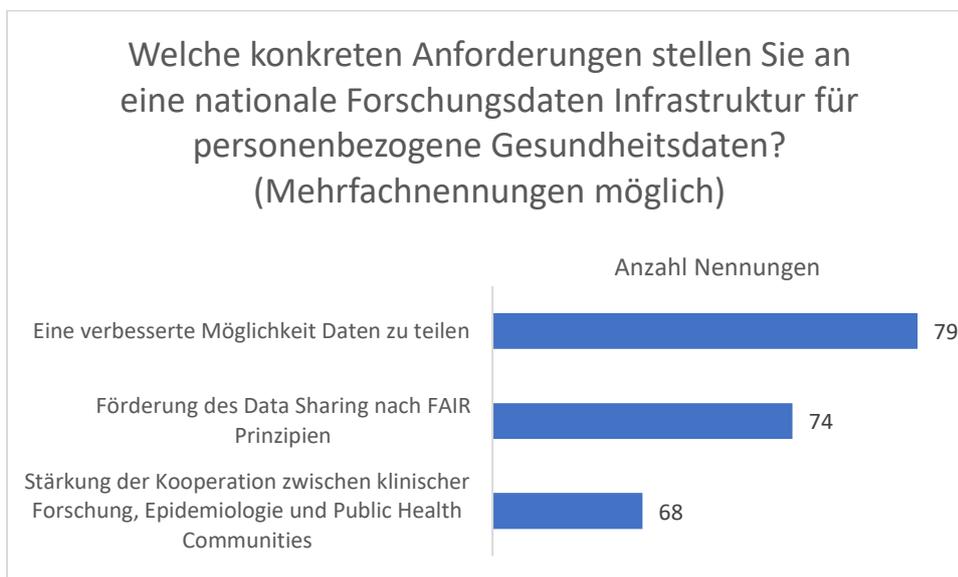


Abbildung 47 „Welche konkreten Anforderungen stellen Sie an eine nationale Forschungsdaten Infrastruktur für personenbezogene Gesundheitsdaten?“ Mehrfachnennungen möglich (n=105, 221 Nennungen)

Hinsichtlich der Beurteilung der Wichtigkeit der avisierten Serviceleistungen einer föderalen Forschungsstruktur werden alle angegebenen Punkte zu einem hohen Maß als wichtig/eher wichtig eingestuft (<65 %). Die größte Wichtigkeit wurden der zentralen Suche nach Erhebungs- und Studiendokumenten (94 %), der zentralen Suche nach Studien und Daten (91 %) sowie dem zentralen Portal zum Upload von Studienmetadaten (79 %) beigemessen.

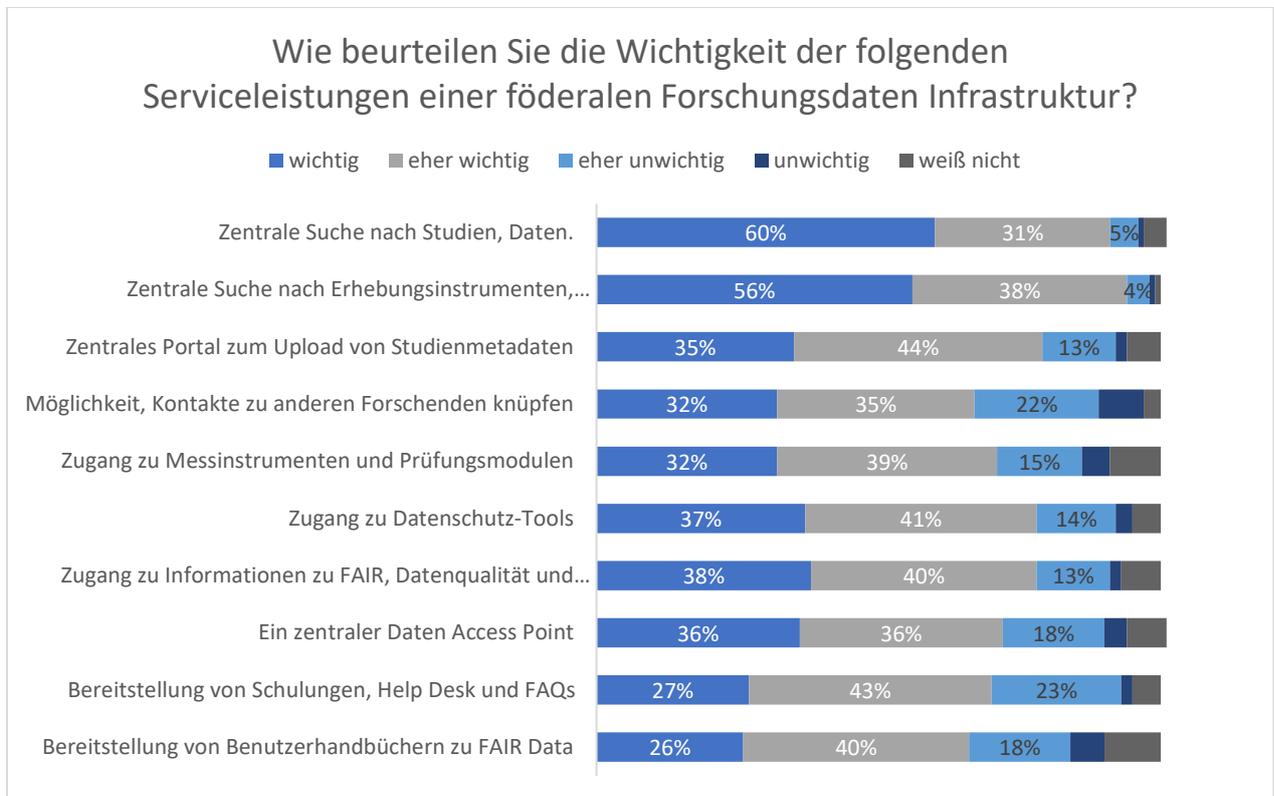


Abbildung 48 „Wie beurteilen Sie die Wichtigkeit der folgenden Serviceleistungen einer föderalen Forschungsdaten Infrastruktur?“ (n=104)

Von jenen, die diese Frage beantworteten, kennen über die Hälfte NFDI4Health nicht.

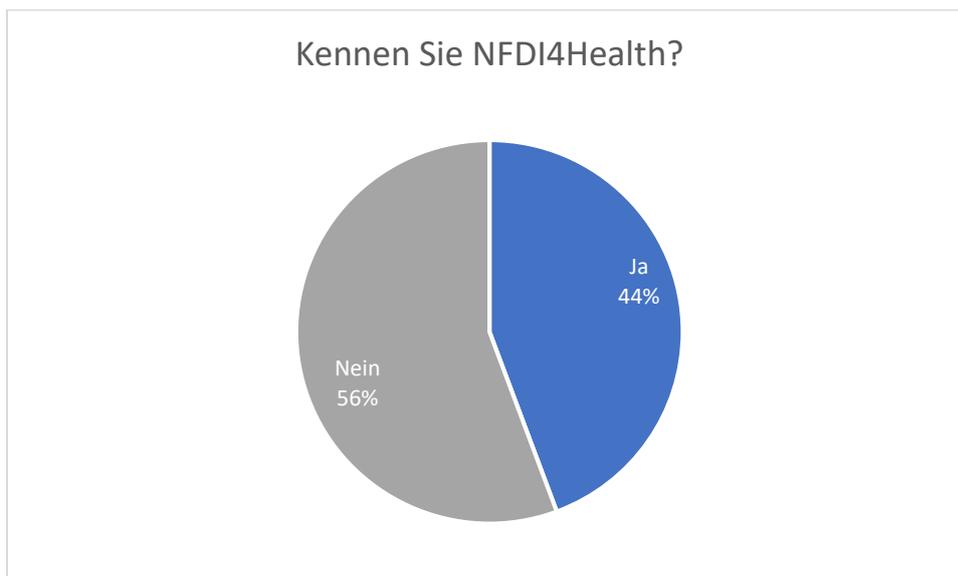


Abbildung 49 „Kennen Sie NFDI4Health?“, (n=106)

Befragte, die NFDI4Health kennen, sind entweder selber am Projekt beteiligt (12 von 27 Nennungen bei offener Frage) oder kennen NFDI4Health aus Vorträgen und/oder Workshops (5 von 27 Nennungen).

Befragte, die NFDI4Health kennen, wurden im Anschluss nach der Bereitschaft, Metadaten auf der NFDI4Health Plattform zu veröffentlichen, gefragt. 85 % können sich dies vorstellen.

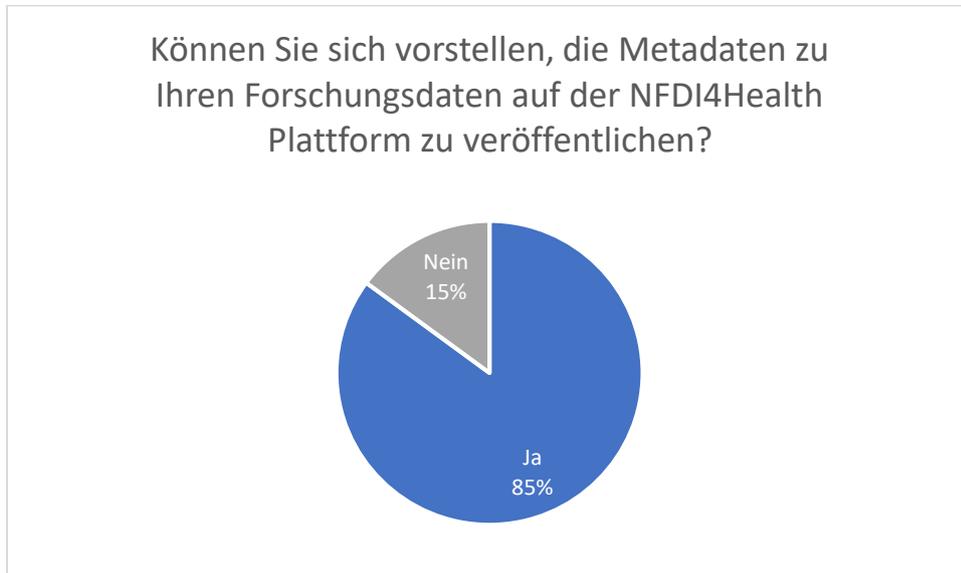


Abbildung 50 „Können Sie sich vorstellen, die Metadaten zu Ihren Forschungsdaten auf der NFDI4Health Plattform zu veröffentlichen?“ (n=47)

Befragte, die NFDI4Health kennen, wünschen sich Informationen zum Projekt primär in Form eines Newsletters (27 Nennungen bei n=47) oder durch Veranstaltungen (25 Nennungen).

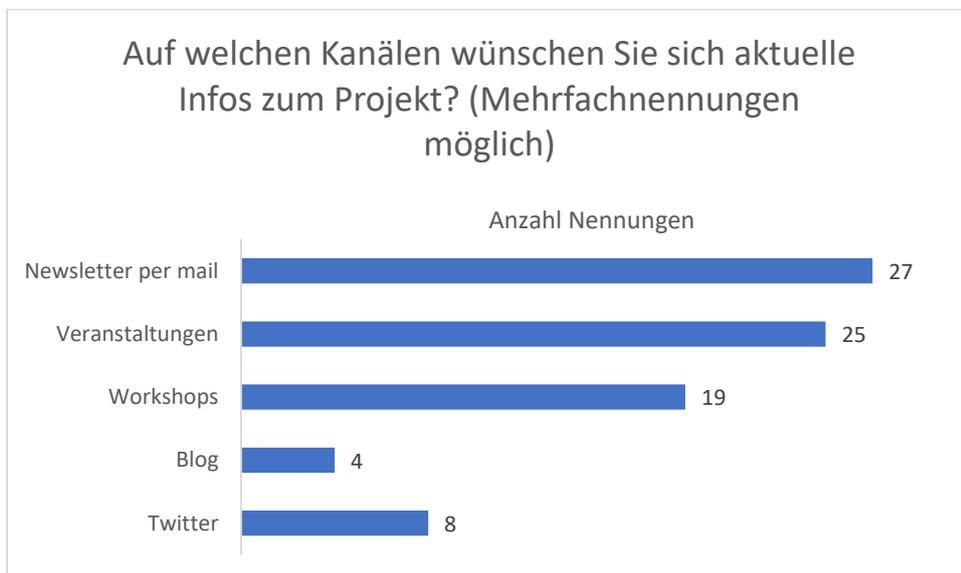


Abbildung 51 „Auf welchen Kanälen wünschen Sie sich aktuelle Infos zum Projekt?“ Mehrfachnennungen möglich (n=47)

Die Bereitschaft, sich an NFDI4Health zu beteiligen, ist am höchsten im Bereich Portaltestung bzw. Feedback zum Portal (9 von n=20) sowie im Bereich Entwürfe und/oder Rückmeldung zu Standards (jeweils 8 Nennungen).

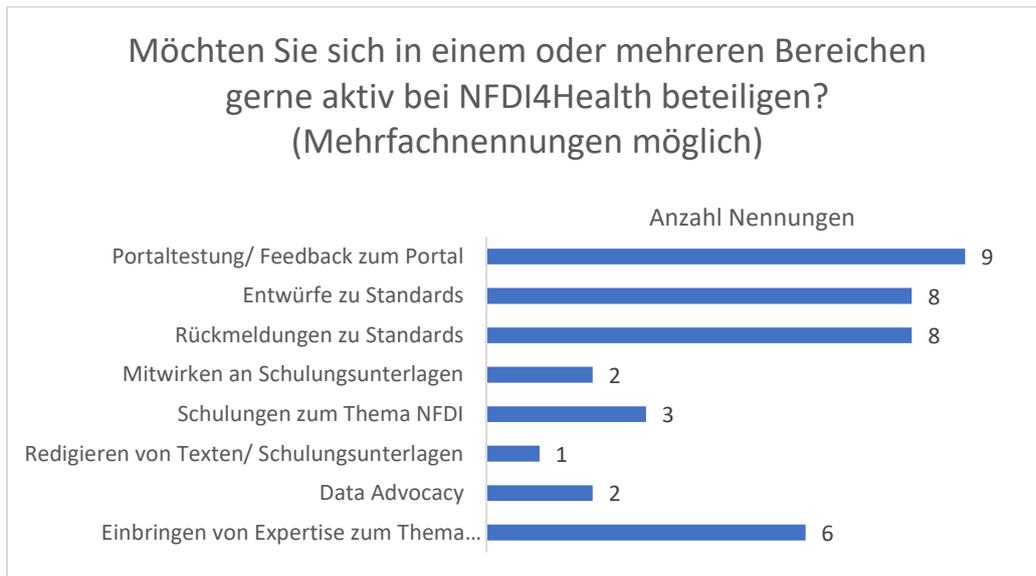


Abbildung 52 "Mögen Sie sich in einem oder mehreren Bereichen gerne aktiv bei NFDI4Health beteiligen? Mehrfachnennungen möglich (n=20)

Die Ergebnisse des vorletzten Fragebogenabschnitts zeigen, dass sowohl die im Antrag identifizierten Anforderungen als auch die konkreten Serviceleistungen von der befragten Community als wichtig und relevant erachtet werden. Ein zentraler Ort für die Suche nach Forschungsdaten und -dokumenten, aber auch für den Upload von Metadaten, sind klare Desiderate, deren sich NFDI4Health annehmen wird.

### Zusammenfassung

Auch die Ergebnisse dieses Fragenkomplexes zeigen, dass NFDI4Health mit seinen Zielen - das Teilen von Daten zu erleichtern, deren Nachnutzung einfacher zu gestalten und somit denjenigen entgegen zu kommen, die bereits Daten nachnutzen – sehr nah an den Bedarfen und Wünschen der Community ausgerichtet ist.

## **Zusammenfassung und Fazit**

Der Survey erreichte überwiegend die Personen, die auch erreicht werden sollten: Forschende in Deutschland, die sich mit den medizinischen Forschungsdaten aus den wichtigsten Kerngebieten befassen. Es gibt gute Gründe für die Annahme, dass diejenigen, die sich an der Befragung beteiligt haben, tendenziell eher einen stärkeren Bezug zum Thema Forschungsdatenmanagement aufweisen als die Grundgesamtheit aller medizinisch Forschenden. Dies ist sicherlich zum Teil der Distribution durch konsortiumsinterne Multiplikator:innen geschuldet.

### **Heterogene Grundgesamtheit aus dem Dunstkreis des NFDI4Health Konsortiums**

Das Gros der Befragten erhebt selber personenbezogene Daten oder verwaltet personenbezogene Daten, die von anderen erhoben wurden. Die Erstellung von Datenmanagementplänen hat sich hierbei noch nicht durchgesetzt. Generell kann nicht von einer Standardisierung im Bereich Forschungsdatenmanagement an den Institutionen bzw. in den Projekten der Befragten gesprochen werden. Was die jeweiligen Standards für die Sicherung, Erfassung, Formatierung und Beschreibung von Forschungsdaten betrifft, herrscht bei der Hälfte der Befragten Unkenntnis über ihr Vorhandensein.

#### **Konkreter Handlungsbedarf I: Etablierung von Standards für die Sicherung, Erfassung, Formatierung und Beschreibung von Forschungsdaten**

Die zum Forschungsdatenmanagement genutzten Tools sind sehr heterogen, trotzdem ist Interoperabilität noch kein Thema. Die Speicherung der Daten erfolgt sowohl während als auch nach dem Forschungsprozess auf lokalen Servern. Mit passenden Repositorien kennen sich die Befragten nicht aus.

#### **Konkreter Handlungsbedarf II: Bewusstsein für die Notwendigkeit von Interoperabilität der FD schärfen**

#### **Konkreter Handlungsbedarf III: Bewusstsein für die überregionale Speicherung der Forschungsdaten stärken**

In der befragten Grundgesamtheit herrscht (vermutlich durch Mechanismen der Selbstselektion) eine große Bereitschaft, Daten mit Dritten zu teilen. Die Meisten haben bereits Daten und Erhebungsinstrumente, meistens mit persönlich bekannten Forschenden/Personen geteilt. Dies geschieht primär über persönliche Kanäle und/oder über persönlich gewährten Zugang. Dies zeigt, dass in dem Bereich noch bislang ungenutzte Potentiale schlummern und die Forschenden hier mehr Informationen und sichere Plattformen für den Austausch benötigen.

#### **Konkreter Handlungsbedarf IV: Schulungen, Guidelines, Informations- und E-Learning-Materialien zum Management personenbezogener Gesundheitsdaten**

Um das Teilen von Forschungsdaten auch außerhalb der eigenen Institution zu möglich zu machen, fordern die Befragten konkrete Nutzungsabsprachen im Vorfeld und/oder eine formale Zitation der Daten. Auch besteht die Bereitschaft, Studienteilnehmenden Daten zur Verfügung zu stellen, allerdings wünschen sich die Teilnehmenden hierzu noch rechtliche Grundlagen bzw. Guidelines und Schulungen. Dies zeigt, dass NFDI4Health mit seinen Zielen, Schulungen zum Thema Forschungsdaten anzubieten, den Wünschen der Community entspricht und dass der Bedarf besteht. Ähnlich verhält es sich mit dem Wunsch nach rechtlichen Grundlagen und Musterformularen für das Teilen von Daten, der ebenfalls durch das Projekt adressiert wird.

#### **Konkreter Handlungsbedarf V: Rechtliche Grundlagen klären, bindende Musterformulare für Nutzungsabsprachen bereitstellen**

Die befragte Grundgesamtheit teilt häufiger eigene Forschungsdaten, als sie Forschungsdaten von Dritten nachnutzt.

Daten von Dritten – primär strukturierte numerische Daten – werden dazu genutzt, einen Überblick über ein Thema zu erhalten oder um die eigene Fragestellung zu schärfen. Dabei begegnen den Forschenden Hürden in Form von ungenügender Beschreibung der Daten, kompliziertem Prozedere bei der Beschaffung und/oder unklare Situationen, was Lizenzrechte bzw. Datenschutz betrifft. Dabei besteht bei der Grundgesamtheit der Befragten durchaus großes Interesse an der Nachnutzung von (Forschungs-)Materialien. Diejenigen, die Daten nachnutzen, suchen sehr gezielt nach ihnen auf spezifischen Projektseiten; es werden also keine Nachweissysteme verwendet, die eine unabhängige Suche ohne Vorwissen zulassen. Hieraus ergibt sich ein weiteres Desiderat an NFDI4Health: einen zentralen Ort für die Suche nach Forschungsdaten zur Verfügung zu stellen.

Im Rahmen des Fragebogens wurden jene Befragten, die keine Daten nachnutzen, nicht gesondert angesprochen. Demnach ist anzunehmen, dass bei dieser Gruppe die Hürden zur Datennachnutzung noch höher sind, was dazu führt, dass Daten nicht nachgenutzt werden, obwohl das Interesse daran vorhanden ist.

### **Konkreter Handlungsbedarf VI: Zentrale Plattform zur gezielten Suche nach Daten zur Nachnutzung**

Wenn man von der naheliegenden These ausgeht, dass diejenigen, welche sich an der Befragung beteiligt haben, tendenziell Forschende sind, die bereits eine höhere Affinität zum Thema Forschungsdatenmanagement haben, werden die tatsächlichen Desiderate im Bereich der medizinischen Forschungsdaten eher noch gravierender sein, als sich aus dieser Befragung ableiten lässt.

## **Ausblick**

Auf Anregung des Konsortiums erfolgt zeitnah eine weitere Versandaktion des Surveys: dieses Mal auch in einer englischen Version, um weitere Personenkreise auch außerhalb des NFDI4Health-Konsortiums zu erreichen. Dadurch besteht die Chance, die Desiderate, die in der ersten Erhebungsrunde zu Tage getreten sind, noch schärfer zu umreißen.

Des Weiteren sollen in weiteren Interviewrunden bzw. Kurz-Surveys die Perspektiven und Desiderate derjenigen, die noch keine Daten nachnutzen, aber auch der Dateninfrastrukturbetreibenden und der Teilnehmenden an klinischen Studien evaluiert werden.

Nach drei Jahren Projektlaufzeit soll der Survey in einer zweiten Runde versendet werden, um zu überprüfen, ob die Desiderate weiterhin im gleichen Umfang bestehen oder ob NFDI4Health bereits einen (positiven) Einfluss zeigen konnte.