

# Forschungsdaten FAIR verwalten

## Methodik und Auswertung der Umfrage

Holger Israel, Esther Tobschall, Frank Tristram

14. Juni 2021

## 1 Abstract

Der Umgang mit Forschungsdaten gewinnt mit fortschreitender Digitalisierung an gesellschaftlicher Aufmerksamkeit. Während ein weitgehender Konsens zu den in den FAIR-Prinzipien dargestellten Anforderungen an gutes Forschungsdatenmanagement besteht, ist ihre Umsetzung eine längerfristige Aufgabe.

Das im Rahmen der *Nationalen Forschungsdateninfrastruktur* (NFDI) gegründete Konsortium NFDI4Phys führte im Frühjahr 2020 eine Umfrage zur Ermittlung des *status quo* und der vordringlichen Bedarfe an das FDM in der Physik durch. In diesem Dokument stellen wir die Methodik und detaillierte Auswertung der Umfrage vor; für den Kontext und unsere Schlüsse verweisen wir auf die Publikation im PhysikJournal.

Wir stellen fest, dass sich die physikalische Forschung durch eine Vielzahl der Methoden und Datenformate auszeichnet. Sie erfährt einen heterogenen Prozess der Digitalisierung, in dem digitale Methoden und eine teils noch analoge Dokumentation nebeneinander stehen. Da wesentliche Teilbereiche der Physik in den bestehenden Konsortien noch nicht vertreten waren, sehen wir den Bedarf eines komplementären Konsortiums NFDI4Phys. Aus der Umfrage ergeben sich als Schwerpunkte unserer Arbeit die Aushandlung von *community standards* für das FDM in der Physik, Angebote zur Steigerung der Datenkompetenz (*data literacy*) sowie die Entwicklung eines modularen Baukastens flexibler FDM-Tools zur Umsetzung der FAIR-Prinzipien.

## 2 Methodik

**Ziel der Umfrage** Mit der Umfrage sollten folgende Ziele erreicht werden:

1. Erfassung und Übersicht des *status quo* des FDM in der Physik
2. Information der Teilnehmer\*innen über die NFDI und Anstoß, sich mit FDM auseinanderzusetzen
3. Priorisierung von Themen und Feedback aus der Community um somit den *bottom-up*-Ansatz der NFDI auch für NFDI4Phys zu verfolgen

4. Antwort auf die Frage, in welchen Teilgebieten der Physik der Bedarf an FDM besonders hoch ist. Damit verbunden: Ansprechen von Gruppen, die an einer Mitgestaltung der NFDI4Phys interessiert sind.

Die aus 27 Fragen bestehende Umfrage wurde Anfang 2020 erarbeitet und gliedert sich in folgende Abschnitte:

- A: Datenschutz
- B: Informationen zur Einrichtung
- C: Engagement in der NFDI
- D: Umgang mit den eigenen Forschungsdaten
- E: Ausblick [im Sinne künftiger Bedürfnisse]

Die Umfrage wurde mit der Open Source-Plattform *LimeSurvey*<sup>1</sup> umgesetzt und verwaltet und an der Technischen Informationsbibliothek Hannover gehostet. Die Umfrage wurde im anonymisierten Modus durchgeführt, wir verzichteten also auf die Möglichkeit, einzelne Zugriffe auf die Umfrage tracken und identifizieren zu können. Die bereitgestellten Rohdaten wurden zusätzlich von Angaben bereinigt, die auf Einzelpersonen hätten schließen lassen.

## 2.1 Bewerbung der Umfrage

Auf die Umfrage wurde auf drei Wegen aufmerksam gemacht: Erstens durch gezielte E-Mails an einen Verteiler aus > 1100 institutionellen und persönlichen Adressen. Der Verteiler basiert auf den in DFG-GEPRIS hinterlegten Kontaktdaten der Antragsteller DFG-geförderter Projekte, ergänzt um Adressen von den Webseiten der Fachbereiche und Institute. Die Adressaten erhielten eine Einladung zur Umfrage mit Erläuterungen zum NFDI-Kontext und zum Anliegen des Konsortiums NFDI4Phys, dessen "genaue Struktur, Ziele und Aufgaben" durch die Umfrage dem Bedarf und dem *bottom-up*-Ansatz der NFDI entsprechend definiert werden sollen. Außerdem wurden die Adressaten ermuntert, die Umfrage in ihren Institutionen zu teilen und weiterzuverbreiten. Durch diese parallelen Kanäle von Funktions- und persönlichen Mailadressen sowie internen Verteilern ist uns die genaue Gesamtzahl der Adressaten nicht bekannt, einige Personen wurden so sicherlich mehrfach kontaktiert.

Zweitens wurde die Umfrage über Flyer beworben, die kurz über Konsortium und Umfrage informierten und mittels eines QR-Codes auf die Umfrage verwiesen. Diese Flyer sollten auf den DPG-Frühjahrstagungen 2020 verteilt werden. Nachdem die Frühjahrstagungen wegen der CoViD19-Pandemie abgesagt worden waren, haben wir die Flyer den Physik-Fachbereichen per Post mit Bitte um Verteilung zugesandt.

Als dritter Verteilungsweg wurde die Umfrage schließlich auf der Webseite des Konsortiums NFDI4Phys verlinkt.

---

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

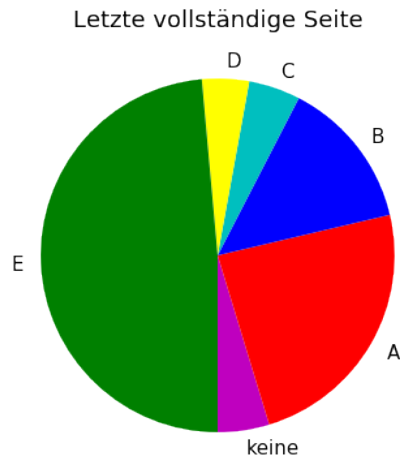


Abbildung 1: Die zuletzt bearbeitete Seite (Abschnitt) der Umfrage lässt auf eine knappe Hälfte an vollständigen Antworten schließen.

## 2.2 Dauer und Verlauf der Umfrage

Die Umfrage wurde am 6. März 2020 mit dem Versenden der Werbemails eröffnet. Ursprünglich sollte die Umfrage bis zum 10. April 2020 laufen. Insbesondere wegen des sich in der Zwischenzeit entwickelnden pandemischen Geschehens beschlossen wir, die Umfrage um einen Monat bis zum 10. Mai 2020 zu verlängern. Dazu wurden am 6. April 2020, einen Monat nach dem ersten Mailing alle Adressaten über die Mailingsliste erneut kontaktiert. Allerdings war zu diesem Zeitpunkt schon eine deutliche Sättigung der Teilnahme zu verzeichnen: Von den 488 Rückläufern gingen 418 (85 %) vor dem 14. April 2020 ein.

## 3 Ergebnisse

Insgesamt wurde die Umfrage 534-mal aufgerufen, davon wurden bei 488 Aufrufen auch Fragen beantwortet, wobei festzustellen ist, dass letztlich in 237 Fällen alle Abschnitte der Umfrage bearbeitet worden sind (vgl. Abb. 1). LimeSurvey notiert die jeweils letzte beantwortete Seite einer aufgerufenen Umfrage.

### 3.1 Fragen 1 und 2: Geographische Verteilung und Institutionen

Die überwiegende Mehrheit von insgesamt 86 % der Antwortenden geben an, an einer Universität, Technischen Universität oder Fachhochschule zu arbeiten, dabei entfallen alleine 71 % der Antworten auf Universitäten. Von außeruniversitären Forschungseinrichtungen stammen 14% der Antworten. Nur wenige Teil-

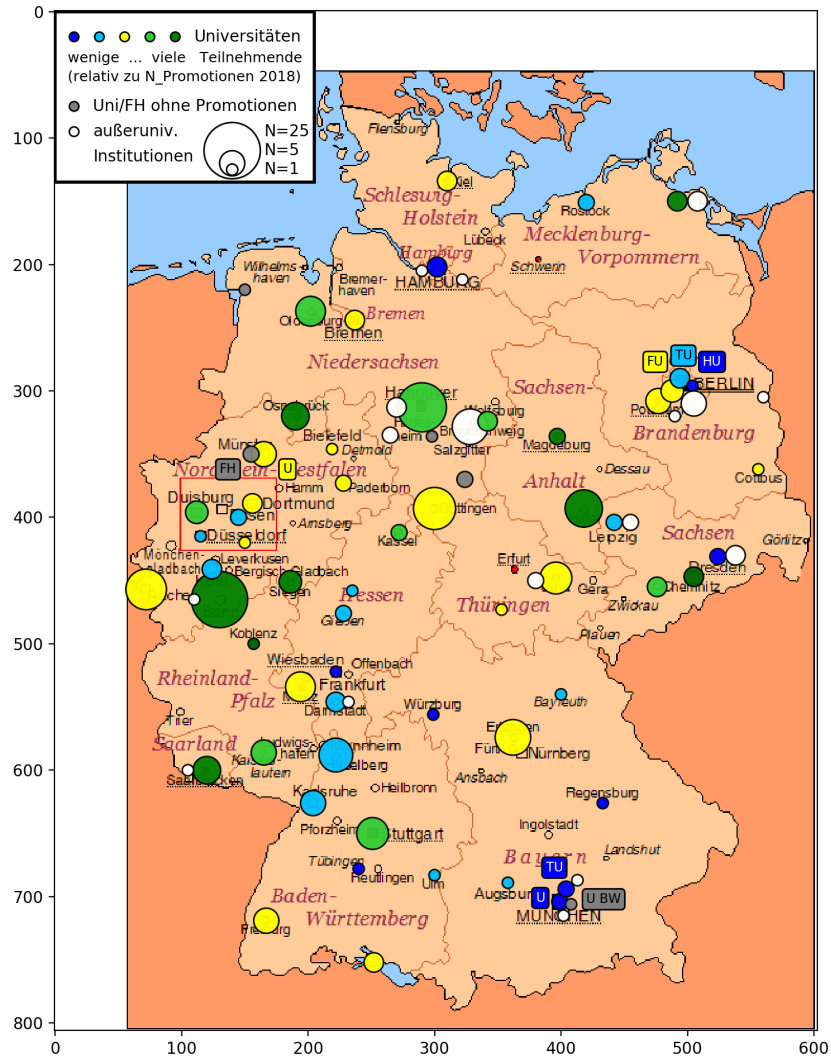


Abbildung 2: Antworten auf Frage 1: “Bei welcher Institution sind Sie tätig?” Die Fläche der Symbole ist proportional zur Anzahl der Antworten. Bei Universitäten gibt die Farbe an, ob relativ zur Anzahl der Promotionen sehr viele (dunkelgrün), viele (hellgrün), eine mittlere Anzahl (gelb), wenige (hellblau) oder sehr wenige (dunkelblau) Rückmeldungen zu verzeichnen sind. Hochschulen ohne Physikpromotionen haben graue Symbole. Außeruniversitäre Einrichtungen sind in weiß dargestellt. Die Abbildung benutzt eine von Korny78 in den Wikimedia Commons unter CC-BY-SA-3.0 veröffentlichten Karte.

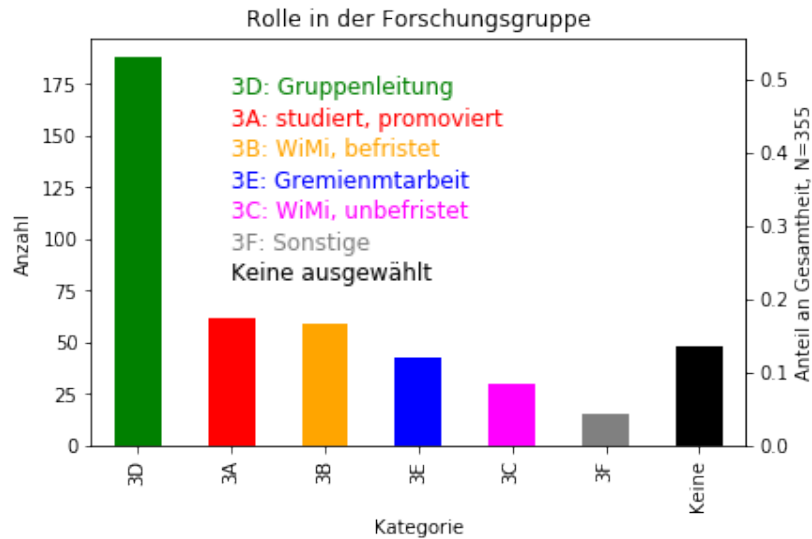


Abbildung 3: Antworten auf Frage 3: “In welcher Rolle sind Sie in Ihrer Forschungsgruppe tätig? (Mehrfachauswahl möglich)

nehmende sind in der Industrie tätig. Bemerkenswert ist, dass außer der Jacobs University Bremen alle deutschen Universitäten mit physikalischer Forschung an der Umfrage teilgenommen haben, wenn sie auch mit teils sehr unterschiedlichem Anteil an Antworten repräsentiert sind. Als Proxy für die schwer zu definierende “Größe” einer Physik-Fakultät bzw. eines Physik-Fachbereichs ziehen wir die Anzahl der abgeschlossenen Promotionen heran, die in den jährlichen “Statistiken zum Physikstudium an den Universitäten in Deutschland” der DPG veröffentlicht werden. [1]. Von den nach dieser Metrik größten Fachbereichen (an den beiden großen Münchner Universitäten sowie in Hamburg) bekamen wir dabei auffällig weniger Rückmeldungen. Innerhalb der verschiedenen Fachbereiche wird die Umfrage in sehr unterschiedlichem Maße weiterverteilt worden sein. Auch ein zusätzlicher “Corona-Bias” in der geographischen Verteilung erscheint plausibel (vgl. Abschnitt 2.1).

### 3.2 Frage 3: Rolle in der Forschungsgruppe

Personen mit Gruppenleiterfunktion stellen eine knappe Mehrheit von 53 % (Abb. 3) der Antwortenden. 17 % der Teilnehmenden ordnen sich der Gruppe der Studierenden und Promovierenden zu, auf die Gruppe der befristet angestellten wissenschaftlichen Mitarbeitenden entfallen ebenfalls 17 %. Unbefristete wissenschaftliche Mitarbeitende sind mit 8 % seltener vertreten. Unabhängig von ihrem sonstigen Status wirken 12 % der Befragten in Gremien mit. Hingegen sehen sich 14 % keiner der vorausgewählten Gruppen zugehörig.



Abbildung 4: Antworten auf Frage 4: “Welchen Fächern laut der DFG-Fachsystematik ordnen Sie die inhaltlichen Themen Ihrer Forschungsgruppe zu? (Mehrfachauswahl möglich)”. Zahlen in Fettdruck geben die Summe der Zuordnungen zum jeweiligen Fach wieder.

### 3.3 Frage 4: Fachverteilung

Die Teilnehmenden der Umfrage konnten die inhaltlichen Themen ihrer Forschung fünfzehn fachrelevanten Disziplinen aus der DFG-Fachsystematik zuordnen (Codes in Abb. 4). Dabei unterscheiden sich die von der DFG definieren Fächer teils auffällig in Granularität und “Größe”. Die Verteilung der Antworten erscheint uns als hinreichend repräsentativ für die Forschung in der Physik. Jeweils 25 % der Teilnehmenden forschen in der “Experimentellen Physik der kondensierten Materie” (307-01) und der “Optik, Quantenoptik, Physik der Atome, Moleküle und Plasmen” (308-01). Diese beiden Fächer finden sich als Kernthemen bei den DPG-Sektionen “Kondensierte Materie” und “Atome, Moleküle, Quantenoptik und Plasmen” wieder. An dritter Stelle folgt die “Kern- und Elementarteilchenphysik, Quantenmechanik, Relativitätstheorie, Felder” (309-01) (vertreten in der DPG-Sektion “Materie und Kosmos”) mit 17 %. Weitere häufig genannte Fächer sind die “Statistische Physik, Weiche Materie, Biologische Physik, Nichtlineare Dynamik” (310-01) und die “Theoretische Physik der kondensierten Materie” (307-02) mit jeweils je 12 %. Aus der “Astrophysik und Astronomie” (311-01) stammen 8 % der Antworten. Insgesamt 20 % der Teilnehmenden sehen sich einem chemienahen Fach (302, 303, 306) angehörig. Wir verzeichnen nur wenige Antworten aus der Geophysik. Dies mag der Tatsache geschuldet sein, dass diese Fächer an Universitäten üblicherweise organisatorisch von der Physik getrennt sind.

### 3.4 Frage 5: Art der Forschungsaktivität

Abbildung 5 zeigt, dass in 80 % der befragten Forschungsgruppen Grundlagenforschung betrieben wird. Mit 39 % hat die angewandte Forschung ebenfalls einen hohen Stellenwert in den Forschungsgruppen. Das Spektrum der Aktivitäten in den Forschungsgruppen wird vervollständigt durch Auftragsforschung

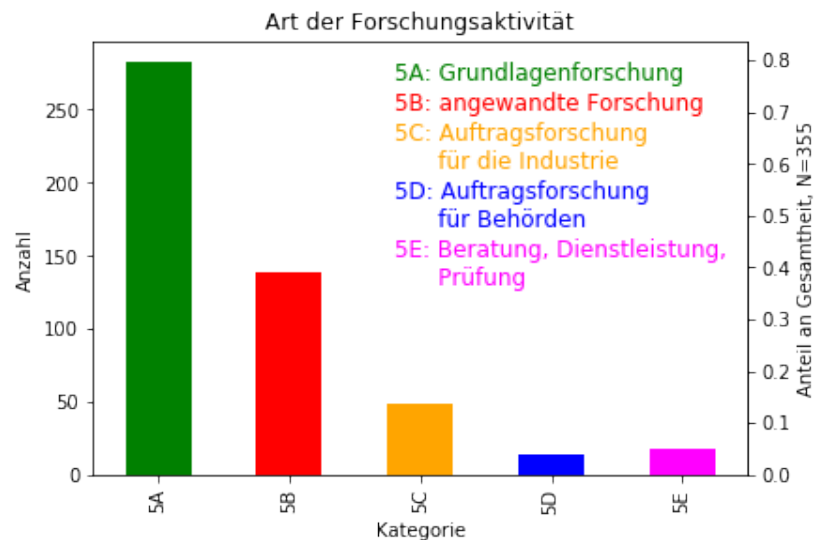


Abbildung 5: Antworten auf Frage 5: “Welche Art von Aktivität wird in Ihrer Forschungsgruppe betrieben? (Mehrfachauswahl möglich)

für die Industrie (14 %) bzw. für Behörden (4 %) sowie Beratung, Dienstleistung und Prüfung (5 %).

### 3.5 Fragen 6 und 7: NFDI

Die Antworten auf Frage 6 “Ist Ihnen die NFDI-Initiative der GWK bekannt?” liefern eine Momentaufnahme aus dem Frühjahr 2020 (siehe Abb. 6, links): Zu diesem Zeitpunkt war einer knappen Mehrheit von 54 % der Antwortenden die Initiative zum Aufbau einer nationalen Forschungsdateninfrastruktur bekannt. Bereits aktiv an der Gestaltung der NFDI beteiligt waren 16 % der Antwortenden (insb. durch Engagement in einem der Konsortien), weitere 16 % sahen sich gut über die NFDI informiert, während sie 22 % zumindest ein Begriff war. Damit kannten zum Zeitpunkt der Erhebung 46 % die NFDI (noch) nicht.

Die sich anschließende Frage 7 ermittelte die Beteiligung der Befragten an geplanten Konsortien und hatte zusätzlich die Funktion, die NFDI-Initiative und die Rolle der Konsortien in dieser zu erläutern: *Im Rahmen der NFDI-Initiative sollen sich etwa 30 Konsortien bilden, die dann gemeinsam eine verteilte Infrastruktur für Forschungsdaten in Deutschland bereitstellen werden. Bei der Gründung der Konsortien finden sich Vertreter von Forschungsinstitutionen zusammen, die Interesse haben, die Entwicklung von Infrastrukturen und Standards für das Forschungsdatenmanagement (FDM) mitzugestalten oder FDM-Dienstleistungen erbringen wollen. Ist Ihre Arbeitsgruppe oder Ihr Fachbereich an einem der genannten Konsortien beteiligt?* Zur (Mehrfach-)Auswahl

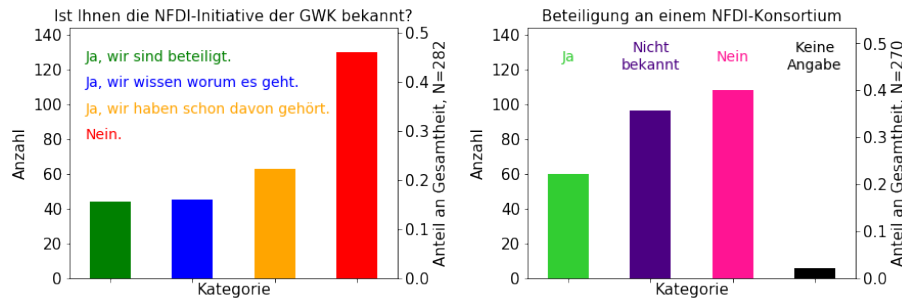


Abbildung 6: *Linke Seite:* Antworten auf Frage 6: Ist Ihnen die NFDI-Initiative der GWK bekannt? *Rechte Seite:* Antworten auf Frage 7: [...] Ist Ihre Arbeitsgruppe oder Ihr Fachbereich an einem der genannten Konsortien beteiligt?

zur Beantwortung der Frage standen 11 Konsortien sowie die Angaben “Nein” und “nicht bekannt”. Von den Antwortenden geben 22 % an, bereits an einem NFDI-Konsortium beteiligt zu sein (siehe Abb. 6, rechts für Erläuterungen zu den Konsortien): Von den zur Auswahl stehenden Konsortien wurde PAHN-PaN (Teilchenphysik) von 21 Personen und damit am häufigsten genannt. Drei weitere Konsortien erhielten jeweils 13 Nennungen: Astro-NFDI, FAIRmat und NFDI4Phys. (Nach Ende der Umfrage haben sich Astro-NFDI und PAHN-PaN für die zweite Antragsrunde zum Konsortium PUNCH4NFDI zusammengetan.) Die weiteren aufgeführten Konsortien DAPHNE, NFDI4MSE, NFDI4Ing, NFDI4Chem und NFDI4Life wurden bis zu viermal genannt. An NFDI Neuroscience und NFDI4Earth war niemand beteiligt. Während 40 % antworten, daß Ihre Arbeitsgruppe oder ihr Fachbereich nicht in der NFDI engagiert sind, geben 36 % an, ihnen sei keine Beteiligung bekannt.

Insgesamt 34 mal registrieren wir ein “Nein” für alle Optionen der Frage 7. Sofern diese Personen bei Frage 6 angegeben haben, an der NFDI beteiligt zu sein (sechs Fälle), interpretieren wir dies so, daß diese Personen keine Angaben dazu machen möchten, in welchem Konsortium sie aktiv sind. Es ist auffällig, daß diese Gruppe wo möglich, meist mit “Nein” antwortet und keine Angaben zu Arbeitsgruppe und Forschungsgebiet macht.

### 3.6 Frage 8: Vertretung durch NFDI-Konsortien

Zielsetzung der sich anschließenden Frage 8 war, herauszufinden, inwieweit Wissenschaftler\*innen den Bedarf für ein - neben u.a. PAHN-PaN, Astro-NFDI, DAPHNE und FAIRMat - weiteres Konsortium für die Physik sehen: *“Die genannten Konsortien haben sich auf Initiative einzelner Fachgemeinschaften zusammengefunden; deren Leistungen werden jedoch für alle Forschenden zur Verfügung stehen (unabhängig von ihrer Beteiligung an einem Konsortium). Um auch die Physik-Gebiete zu beteiligen, die noch nicht durch ein anderes der oben genannten Konsortien vertreten werden, planen wir das Konsortium NF-*



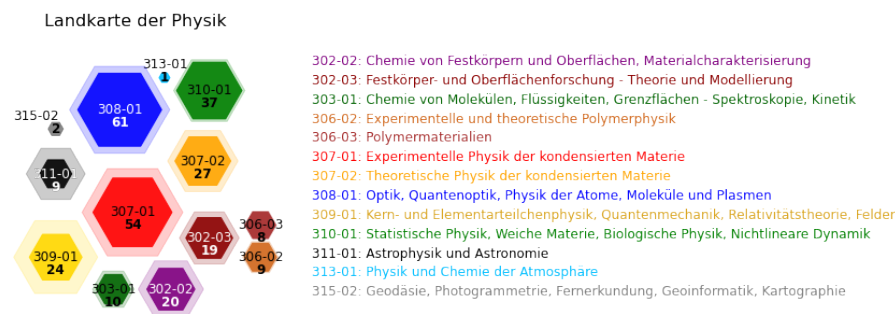


Abbildung 7: Antworten auf Frage 8: “[...] Sehen Sie Ihre Belange bezüglich Forschungsdaten bereits durch ein anderes Konsortium hinreichend vertreten?”. Wie Abb. 4, aber mit inneren Waben proportional zur Anzahl derer, die sich noch nicht durch ein NFDI-Konsortium vertreten sehen.

*DI4Phys.* Sehen Sie Ihre Belange bezüglich Forschungsdaten bereits durch ein anderes Konsortium hinreichend vertreten?” Damit ist sie explizit auf die Ermittlung des Potentials für NFDI4Phys gemünzt.

Insgesamt geben 43 % der die Frage 8 Beantwortenden an, dass sie sich bereits durch eines der anderen Konsortien vertreten fühlen. Die Verteilung der genannten Konsortium unterscheidet sich nur wenig von der vorigen Frage. Demgegenüber stehen 57 %, die ihre Belange noch durch kein Konsortium vertreten sehen. Abbildung 7 wiederholt die “Landkarte der Physik” aus Abb. 4 mit abgeschwächten Farben. Die Fläche der inneren Waben ist proportional zur Anzahl derjenigen Antwortenden, die sich noch durch kein NFDI-Konsortium vertreten sehen.

Wir beobachten eine auffälligen Verteilung auf die DFG-Fächer: Unter den häufig (>25 mal) genannten Fächern war der Grad der Vertretung durch NFDI-Konsortien im Frühjahr 2020 mit 16 % im Fach 310 (Statistische Physik, Weiche Materie, Biologische Physik, Nichtlineare Dynamik) am geringsten. Einen ähnlich niedrigen “Versorgungsgrad” von 17 % und in absoluten Zahlen die meisten noch nicht vertretenen Gruppen finden wir im Fach 308 (Optik, Quantenoptik, Physik der Atome, Moleküle und Plasmen). Diese beiden Fächer haben sich daher als Schwerpunkte für NFDI4Phys herausgebildet. Dennoch stellen wir fest, dass es auch in der Festkörperphysik eine wesentliche Quote “NFDI-unversorgter” Gruppen gab und vermutlich gibt (Versorgungsgrad von 30 % über die Fächer 307-01 und 307-02 gemittelt). Nur in der Astronomie und Astrophysik (Fach 311, 64 %) und der Kern- und Elementarteilchenphysik mit Quantenmechanik, Relativitätstheorie und Feldern (Fach 309, 52 %) sieht sich bereits je eine Mehrheit in ihren FDM-Belangen durch ein NFDI-Konsortium vertreten.

Diese Zahlen lassen die Schlußfolgerung zu, dass die Notwendigkeit für ein weiteres Konsortium im Spektrum der physikrelevanten NFDI-Initiativen besteht und damit das Potential für die Etablierung eines Konsortiums NF-

DI4Phys gegeben ist. Ebenso wie bei den beiden vorigen Fragen stellen die Antworten auf Frage 8 der Umfrage einen Momentaufnahme zum Zeitpunkt der Umfrage dar.

### **3.7 Fragen 9 und 10: Mitarbeit in NFDI4Phys und Kontaktdaten**

Die Fragen 9 und 10 dienen der Identifizierung weiterer potentiellen Partnern für das Konsortium NFDI4Phys. Dabei äußern 26 % derjenigen, die diese Frage beantwortet haben, Interesse an einer Mitarbeit. Weitere 26 % möchten über den Fortgang der Entwicklung informiert bleiben, während 48 % angeben, keine Zeit bzw. kein Interesse zu haben.

An diesem Punkt in der Umfrage konnte eine Mailadresse zur Kontaktaufnahme angegeben werden. Davon machten 129 Personen Gebrauch, mehr als die Hälfte derer, die auf Frage 9 antworteten. Das sich formierende Konsortium NFDI4Phys konnte auf diese Weise den Kreis der den Antrag mittragenden Institutionen und der teilnehmenden Forschenden wesentlich und erfolgreich vergrößern.

### **3.8 Frage 11: Erzeugte Datenformate**

Wie Abbildung 8 zeigt, werden Forschungsdaten in der Physik in einer Vielzahl an (Klassen von) Datenformaten erzeugt. Dabei sind Textformate (wie .csv, .txt, .asc; bei 84 % der Antwortenden), Bilder in Raster- oder Vektorformaten (bei 76 %) und Software-Code (bei 71 %) am weitesten verbreitet. Modellierungen (z.B. CAD), Spreadsheets und Video- oder Audiodateien werden bei je etwa einem Drittel der Antwortenden zur Aufzeichnung von Forschungsdaten genutzt. Seltener werden Datenbanken als erzeugte Datenformate berichtet. Unter “Sonstiges” wird als häufigste Kategorie auf Binärformate, insb. `hdf5`, aber auch `fits` und proprietäre Geräteformate, verwiesen.

### **3.9 Frage 12: Datenvolumen**

Die Frage nach dem jährlich anfallenden Datenvolumen ist so formuliert, dass sie abfragt, wieviel “zur Einhaltung der guten wissenschaftlichen Praxis archiviert werden soll/muß”. Hier gibt es also einen Ermessensspielraum, welche Rohdaten und Zwischenergebnisse in diese für Nachnutzung relevante Kategorie fallen. Abbildung 9 veranschaulicht die Physik als in großen Teilen datenintensive Wissenschaft: In 23% der Gruppen müssen pro Jahr > 10 TB gespeichert werden, in weiteren 31% zwischen 1 TB und 10 TB.

### **3.10 Frage 13: Modi der Datenerzeugung**

Entsprechend der vielfältigen Arbeitsabläufe in der physikalischen Forschung zeigt die Umfrage, dass in den befragten Forschungsgruppen Forschungsdaten

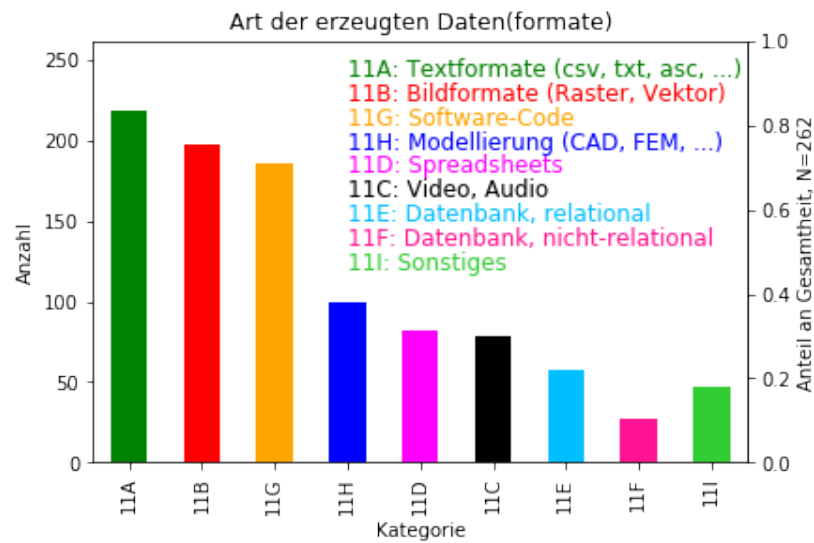


Abbildung 8: Antworten auf Frage 11: “Welche Arten von Daten bzw. Datenformate werden in Ihrer Forschungsgruppe erzeugt? (Mehrfachauswahl möglich)”

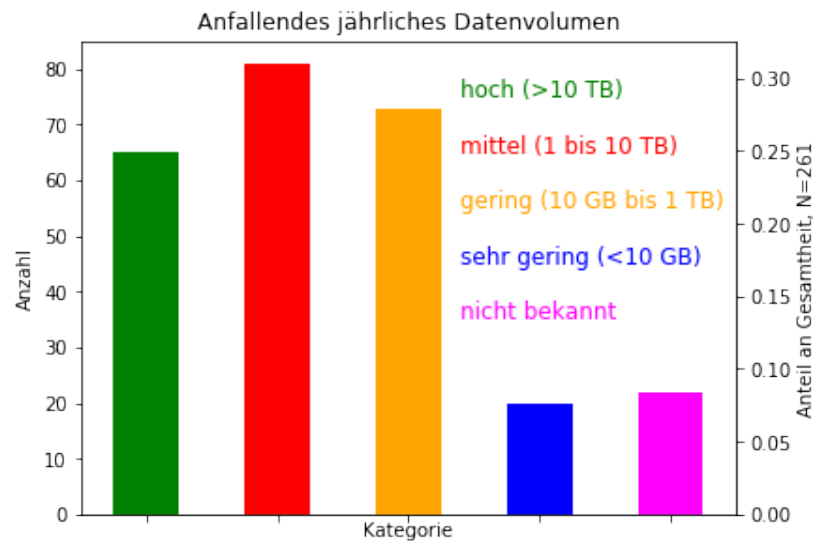


Abbildung 9: Antworten auf Frage 12: “Wie groß ist das in Ihrer Forschungsgruppe anfallende Datenvolumen pro Jahr, das zur Einhaltung der guten wissenschaftlichen Praxis archiviert werden soll/muss?”

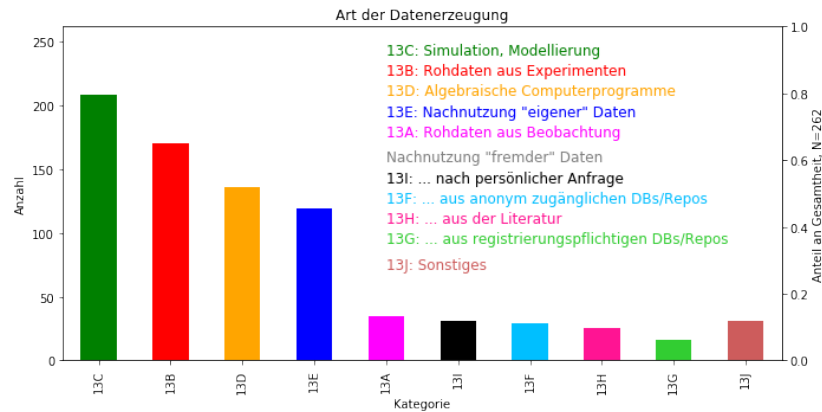


Abbildung 10: Antworten auf Frage 13: “Wie werden die Daten innerhalb Ihrer Forschungsgruppe erzeugt? (Mehrfachauswahl möglich)”

über mehrere Wege entstehen: So werden in 79 % der befragten Gruppen Simulationen oder Modellierung zur Datengewinnung eingesetzt, in 65 % der Gruppen werden Rohdaten durch Experimente aufgenommen. Kombiniert werden Experiment und Simulation von 52 %. Algebraische Computerprogramme kommen ebenfalls bei 52 % zum Einsatz. Nur 45 % der Gruppen nutzen eigene Forschungsdaten z.B. für weitere Forschungsprojekte nach. Noch deutlich seltener werden außerhalb der eigenen Gruppe erzeugte Daten nachgenutzt. Das könnte einerseits bedeuten, dass die Nachnutzung von Daten wissenschaftlich nicht relevant ist, kann aber auch ein Indiz dafür sein, dass die Nachnutzbarkeit von Daten durch äußere Restriktionen (z.B. lückenhafte Dokumentation, proprietäre Datenformate) erschwert wird und dadurch wenig attraktiv erscheint.

### 3.11 Frage 14: Software zur Datenauswertung

Die Antworten zur für Datenprozessierung und -auswertung eingesetzten Software in Abb. 11 zeigen, dass in den meisten Arbeitsgruppen mehrere Softwaretypen parallel verwendet werden. Kommerzielle Software (61 %), Open Source - Software (63 %), selbstentwickelte Software (61 %) und Programmierskripte (62 %) werden alle in ähnlichem Umfang benutzt. Spreadsheets werden nur von einer Minderheit der Gruppen zur *Auswertung* von Daten eingesetzt.

Über Freitextantworten konnte das spezifische Produkt konkretisiert werden. Bei Spreadsheets wird **Microsoft Excel** mit Abstand am häufigsten genannt. Bei kommerzieller wissenschaftlicher Software kommt **Matlab** auf 46 Nennungen, gefolgt von **Origin** mit 29 und **Mathematica** mit 28 Nennungen. Für selbstgeschriebene Skripte ist **Python** (129 Nennungen) am beliebtesten, gefolgt von **C/C++** (88) und, mit weitem Abstand, **Matlab**, **Fortran**, **LabView**, **Mathematica** und einer Vielzahl weiterer Sprachen. Auch bei der Open-Source-Software (wo

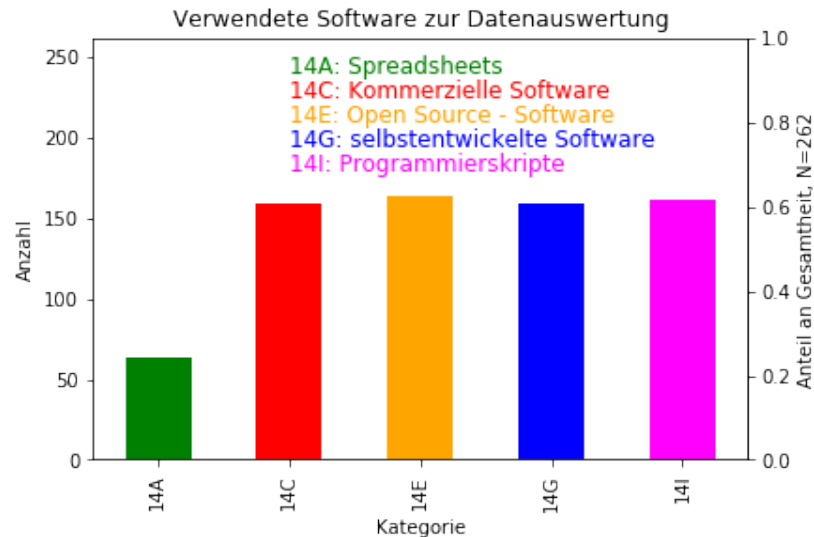


Abbildung 11: Antworten auf Frage 14: “Welche Software wird in Ihrer Forschungsgruppe zur Datenprozessierung und -auswertung verwendet?”

angegeben), dominiert Python deutlich.

### 3.12 Frage 15: Relevante Metadaten im Forschungsfeld

Die Frage nach den für das Forschungsgebiet relevanten Metadaten wurde analog zur Frage nach dem Datenvolumen so formuliert, dass wir nach einem wünschenswerten Soll-Zustand fragten.

Es halten 67 % der Antwortenden administrative Informationen, wie z.B. Zeitstempel oder Name der Datenerhebung relevant, um die Datensätze effizient klassifizieren, dokumentieren und suchen (d.h. auch wiederfinden) zu können. Für 65 % spielen Angaben zur Datenengenerierung eine Rolle, 54 % halten Angaben zum Versuchsobjekt für notwendig. Genau 50 % denken bei relevanten Metadaten an Angaben zu Größen, Einheiten, Wertebereichen, Unsicherheiten. Und 21 % scheinen einen Bedarf für sonstige Metadaten zu sehen, die durch die vorgegebenen Antworten nicht berücksichtigt wurden; bei den Freitextangaben finden sich Punkte wie die Angaben zur Kalibrierung, Dokumentation des Versuchsaufbaus oder die benutzte Software.

Aus Sicht eines strukturierten Forschungsdatenmanagements sind alle diese Metadatentypen notwendig oder zumindest höchst wünschenswert. (In der Regel, aber es mag abhängig vom Forschungsfeld Ausnahmen geben.) Tatsächlich stehen den 24 (9 %) Antworten, die alle vier explizit abgefragten Metadatentypen als wichtig erachten, andere 24 Antwortenden gegenüber, denen keine dieser Metadaten relevant erscheinen. Von Letzteren verweisen allerdings fünf auf nicht

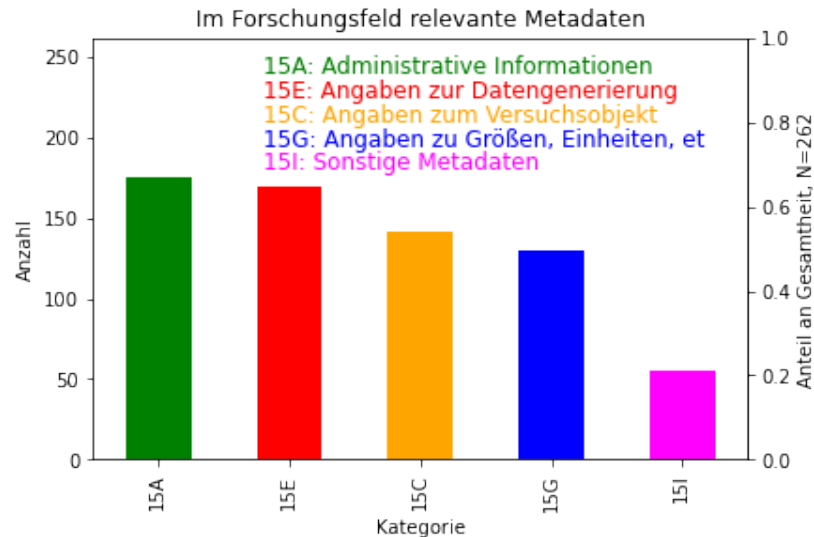


Abbildung 12: Antworten auf Frage 15: “Welche Metadaten sind in Ihrem Forschungsbereich relevant, um Datensätze effizient zu klassifizieren, zu dokumentieren, zu suchen (Mehrfachauswahl möglich)”.

genauer ausgeführte sonstige Metadaten.

Wir schließen daraus einen Bedarf an Aufklärungsarbeit und Bewusstseinsbildung in Sachen FDM.

### 3.13 Frage 16: Dokumentation von Experimenten

Der *status quo* der Dokumentation von Experimenten und damit der Erfassung von Metadaten ist in Abb. 13 dargestellt. Wir stellen fest, dass in der Mehrheit der Gruppen die Dokumentation noch teilweise analog durch persönliche Papierlaborbücher (bei 50 %) bzw. an Geräte gebundene analoge Laborbücher (31 %; Mehrfachauswahl möglich) erfolgt. Ebenso bemerkenswert ist, dass 21 % ein eigenes elektronisches Laborbuchsystem entwickelt und somit vermutlich signifikante Ressourcen investiert haben. Eigenentwicklungen sind sogar häufiger als die Nutzung kommerzieller elektronischer Laborbücher (6 %) und deren Open-Source-Pendants wie eLabFTW (15 %). Eine Definition elektronischer Laborbücher hatten wir nicht vorgegeben, grenzen sie jedoch von einfacheren Log-Dateien ab, welche mit 44 % die zweitverbreiteteste Option sind.

### 3.14 Frage 17: Verwendete Metadatenschemata

Inwieweit erfasste Metadaten zu einer Nutzbarkeit im Sinne der FAIR-Prinzipien beitragen, hängt wesentlich von den benutzten Metadaten-Schemata

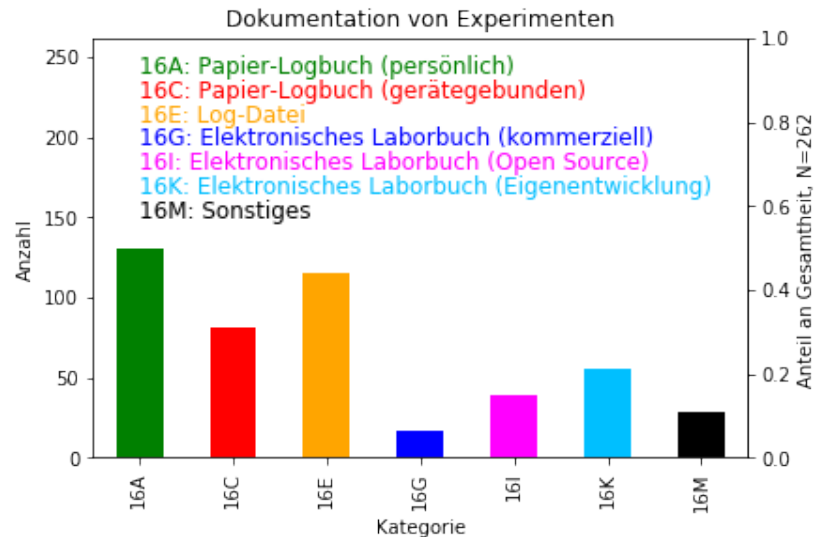


Abbildung 13: Antworten auf Frage 16: “Womit wird der Verlauf eines Experiments dokumentiert? (Mehrfachauswahl möglich)”

ab. Der überwiegende Teil der befragten Gruppen greift hierbei auf selbst-entwickelte Schemata zurück. Standardisierte Schemata werden von zusammen 38 % der Antwortenden genutzt. Speicherung als Textformat überwiegt gegenüber Binärformaten. Mit Abstand am häufigsten werden Metadaten per Hand als Text nach einem eigenen Schema verzeichnet (60 % der antwortenden Gruppen, Mehrfachnennung möglich), vor automatischer Erfassung als Text nach einem eigenen Schema (bei 36 %). Bei 13 % der Arbeitsgruppen findet in der Regel keine Metadatenerfassung statt.

### 3.15 Frage 18: Datenveröffentlichungen

Legt man die 262 Antwortenden zugrunde, die diese Frage bearbeitet haben, so haben 56 % bereits in irgendeiner Form Forschungsdaten veröffentlicht. Dem gegenüber stehen 29 %, die eine Datenveröffentlichung explizit verneinen. Weitere 11 % sind sich der Antwort nicht bewußt.

Drei der “nein” Antwortenden kommentieren: “Das wollen wir aber ändern”, “nur auf Anfrage” und “bisher nur auf Anfrage”. Interessant sind auch in Abb. 15 unter “Sonstiges” zusammengefaßte Antworten, bei denen keine der wählbaren Optionen bejaht, aber ein Freitextkommentar hinterlassen wurde: Während sieben Antworten “Eine Datenveröffentlichung wird erwartet/empfohlen” lauten, heißt es in einem Fall “Eine Datenveröffentlichung wird kritisch gesehen”. Dass alle diese Personen zwar die Umfrage weitgehend bearbeitet haben, jedoch auf Antworten zu Frage 1 nach der *Affiliation* verzichteten, legt nahe,

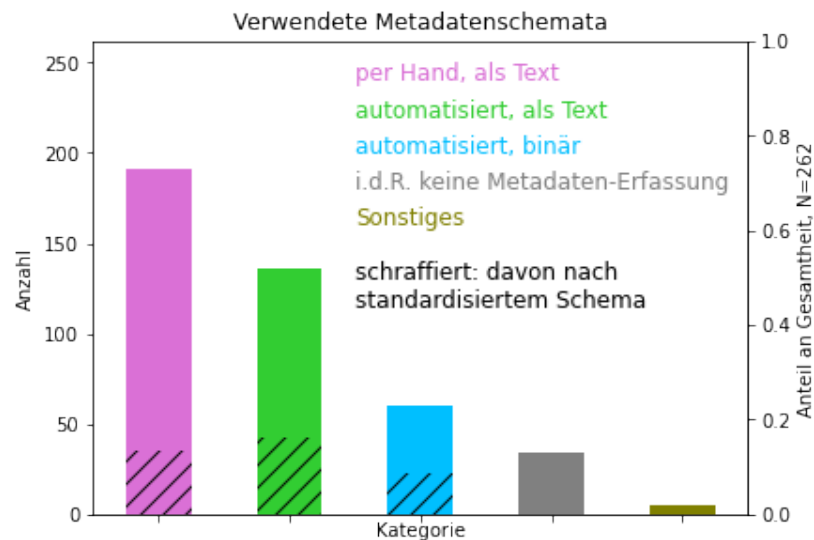


Abbildung 14: Antworten auf Frage 17: “In welcher Form werden Metadaten in Ihrer Forschungsgruppe erfaßt?” (Mehrfachauswahl möglich)

dass sie bewußt nur den “Soll”, nicht den “Ist”-Zustand berichteten.

Die weitaus häufigste Form der Datenpublikation ist in Verbindung mit einer Artikelpublikation (z.B. als Anhang; bei 45 % der insgesamt Antwortenden). Eigenständige Datenpublikationen sind rar (bei 1,5 %). Projekteigene (bei 12 %), von Institutionen bereitgestellte (bei 10 %) und fachspezifische Repositorien (bei 9 %) erfreuen sich ungefähr ähnlicher Beliebtheit. Insgesamt haben bereits 27 % der Befragten ein Datenjournal oder ein Repository für eine Datenpublikation benutzt.

### 3.16 Frage 19: Vorgaben zu Datenveröffentlichungen

Die folgende Frage befasst sich explizit mit den institutionellen Vorgaben zur Veröffentlichung von Forschungsdaten. Für 43 % der Gruppen liegt die Entscheidung darüber, Forschungsdaten zu veröffentlichen oder nicht, bei den Forschenden selber (siehe Abbildung 16). Bei 12 % erwartet ihre Einrichtung eine Datenveröffentlichung; weiteren 17 % wird sie empfohlen. Dem gegenüber stehen vier Antworten (1,6 %), die angeben, eine Datenveröffentlichung werde kritisch gesehen. Mit 26 % ist der Anteil derer groß, die nicht wissen, wie ihre Institution zur Publikation von Forschungsdaten steht.



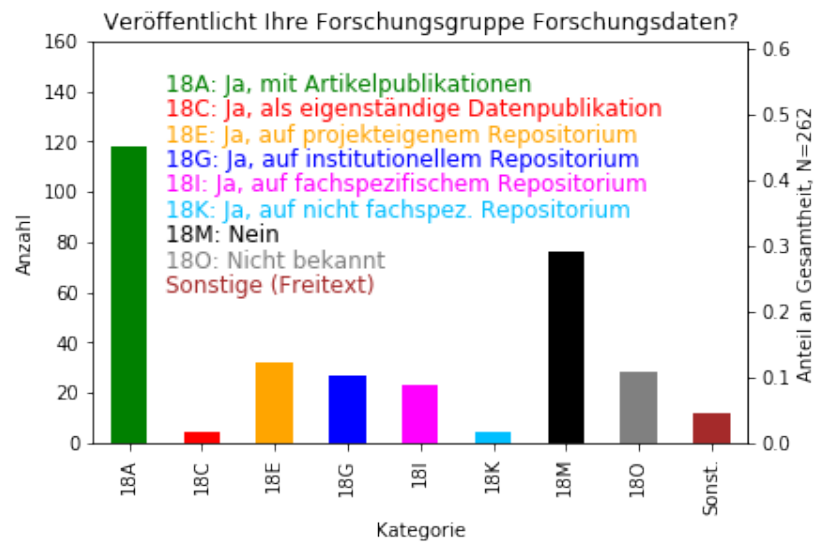


Abbildung 15: Antworten auf Frage 18: “Veröffentlicht Ihre Forschungsgruppe Forschungsdaten?” (Mehrfachauswahl möglich)

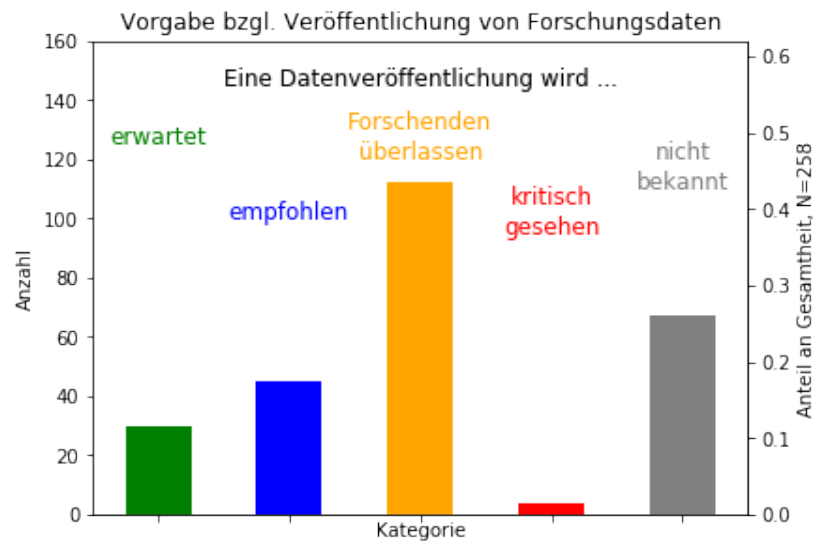


Abbildung 16: Antworten auf Frage 19: “Wie ist die Vorgabe Ihrer Institution bezüglich der Veröffentlichung von Forschungsdaten?” (Mehrfachauswahl möglich)

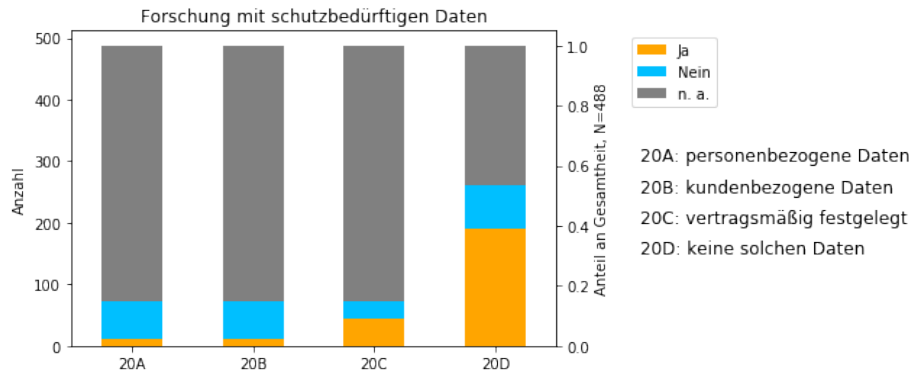


Abbildung 17: Antworten auf Frage 20: “Werden in Ihrer Forschungsgruppe auch schutzbedürftige Daten zu Forschungszwecken verarbeitet?” (Mehrfachauswahl möglich)

### 3.17 Frage 20: Forschung mit schutzbedürftigen Daten

Die Frage, ob zu Forschungszwecken auch schutzbedürftige Daten verarbeitet werden, ist relevant für die Ausgestaltung einer Forschungsdateninfrastruktur. In der medizinischen und medizinnahen Forschung (z.B. bildgebende Verfahren) hat man hier ganz andere Anforderungen als z.B. in der Astrophysik.

Da nur wenige Teilnehmende die ersten drei Teilfragen beantworteten, beziehen wir uns hier zur besseren Vergleichbarkeit auf alle 488 Rückläufer. Von diesen geben 42 % an, mit keinerlei schutzbedürftigen Daten zu arbeiten. Rund 15 % geben an, in ihrer Forschung mit schutzbedürftigen Daten zu tun zu haben: Am häufigsten werden Daten durch Verträge, z.B. mit Kooperationspartnern oder in der Auftragsforschung als schutzwürdig definiert (9 % der gesamten Antworten). Personen- oder kundenbezogene Daten werden von jeweils 2 % der Gruppen verarbeitet. Eine recht große Anzahl von 19 Antwortenden (4 % der Gesamtheit) gibt an, schutzbedürftige Daten zu verwenden, verneint aber die drei abgefragten Optionen. Hier können sowohl andere Arten zu schützender Daten als auch Fehler beim Ausfüllen der Umfrage vorliegen.

### 3.18 Frage 21: Wer darf Daten nachnutzen

Nachnutzung von Forschungsdaten definieren wir als Nutzung nach Ende des ursprünglichen Projekts, d.h. nach Ablauf eventueller Embargofristen. Mitglieder der eigenen Forschungsgruppe haben dieses Recht in 70 % der befragten Gruppen; Mitarbeitende der eigenen Institution immerhin noch bei 32 %. Externe Projektpartner dürfen die Daten von 46 % der befragten Gruppen nachnutzen, müssen sich aber ggf. anmelden. Bei 24 % der Gruppen dürfen alle anderen Forschenden nach Anmeldung auf die Daten zugreifen. Open Access, also Datennachnutzung ohne Zugangskontrolle gestatten 17 % der Forschungsgruppen.

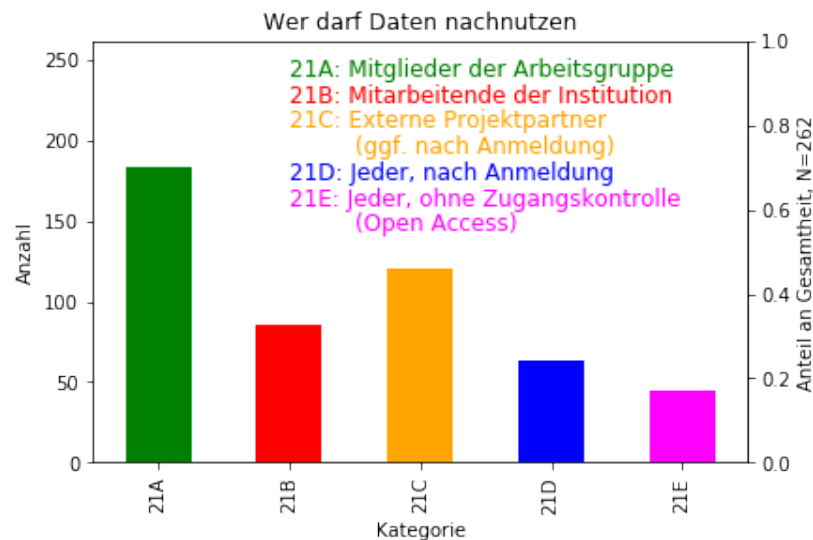


Abbildung 18: Antworten auf Frage 21: “Wer darf Ihre Forschungsdaten nach Ende eines Projekts nachnutzen?” (Mehrfachauswahl möglich)

### 3.19 Frage 22: Qualitätssicherung

Auf die Freitext-Frage nach Kriterien der Datenqualität und Maßnahmen zur Qualitätssicherung sammelten wir 101 Antworten unterschiedlichen Umfangs. Diese Antworten wurden in der Gesamtschau in verschiedene Themen gruppiert, wobei einzelne Antworten bis zu sechs verschiedene Themen ansprechen. Gemäß der Fragestellung können Themen sowohl Kriterien als auch Maßnahmen umfassen. Am häufigsten wurden die Themen “Reproduzierbarkeit” (in 38 Antworten, oft als alleinstehendes Schlagwort gegeben), “Standards, gute wissenschaftliche Praxis” (in 27 Antworten, z.B. der Verweis auf die einschlägigen DFG-Leitlinien) und “statistische Methoden” (in 23 Antworten, z.B. “Bewertung statistischer und vor allem systematischer Fehlerquellen”) angegeben. Aufschlussreich sind auch die folgenden Ränge, deren Achsen “Peer Review”, “Kommunikation” (z.B. “Absprachen in der Arbeitsgruppe” oder “Laufende Diskussion der Daten”), “Dokumentation und Metadaten”, “Software-Tests” und “Vergleich experimentell und numerisch gewonnener Ergebnisse” einen hochdimensionalen Raum der physikalischen Forschung als wissenschaftliches, gesellschaftliches und technisches Unterfangen aufspannen. Die Antworten verweisen darauf, daß die Beschreibung und Wahrung der Qualität von Forschungsdaten ein zentrales Aufgabenfeld der NFDI sein werden, wenn sie den in sie gesetzten Erwartungen gerecht werden möchte.

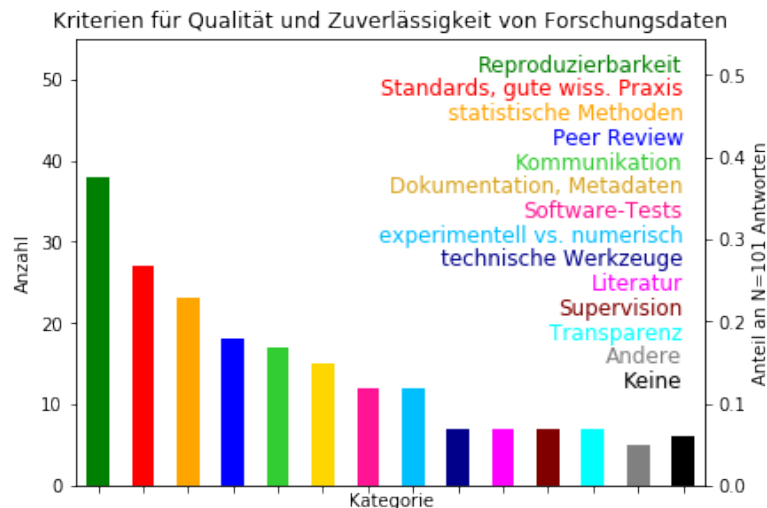


Abbildung 19: Antworten auf Frage 22: “Welche Kriterien wenden Sie an, um die Qualität und Zuverlässigkeit Ihrer Forschungsdaten zu belegen? Welche Maßnahmen werden dafür eingesetzt?” Die Freitext-Antworten wurden kategorisiert, wobei eine Antwort mehrere Kategorien erwähnen kann.

### 3.20 Frage 23: Forschungsdatenmanagement

Wir sehen “strukturiertes Forschungsdatenmanagement” (FDM) als die durch “hausinterne oder von den Forschungsförderern verabschiedete Leitlinien” (*research data management policies*) geregelte Verwaltung, Dokumentierung und Aufbewahrung von Forschungsdaten an.

**Insgesamt 15 % der befragten Gruppen geben an, strukturiertes FDM einzusetzen und die Vorteile zu erkennen.** Weiteren 2,5 % ist es vorgeschrieben, aber sie sehen keine Vorteile darin (siehe Abb. 20).

Zusammen 36 % der befragten Gruppen benutzen momentan kein strukturiertes FDM, und geben dafür einen oder mehrere Gründe an: Am häufigsten wird “nicht hinreichender Support” genannt (22 %), eng gefolgt von zu hohem Aufwand (20 %). Für 10 % sind keine Vorteile ersichtlich.

Der Diskurs um strukturiertes FDM ist bei weitem noch nicht in der gesamten Community angekommen: In insgesamt 38 % der befragten Arbeitsgruppen ist dies kein Thema. Eine nicht kleine Gruppe von 16 % der Befragten wählte explizit die Option “keine Antwort”, möchte also bewußt keine Auskunft geben.

### 3.21 Frage 24: Erhöhung der Akzeptanz von FDM

Im Anschluss fragten wir, was die Akzeptanz von (strukturiertem) FDM erhöhen könnte. Fünf Optionen waren als Vorauswahl gegeben, zusätzlich stand ein Frei-

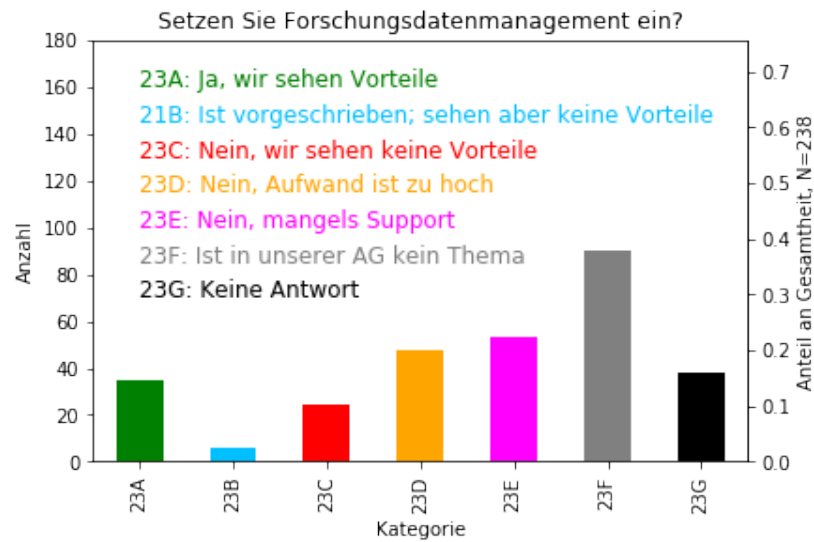


Abbildung 20: Antworten auf Frage 23: “Der Umgang mit Forschungsdaten wird immer häufiger durch hausinterne oder von den Forschungsförderern verabschiedete Leitlinien geregelt. Parallel dazu bieten Forschungsinstitutionen oder externe Einrichtungen Infrastrukturen für das strukturierte Forschungsdatenmanagement (FDM), also für die Verwaltung, Dokumentierung und Aufbewahrung der Forschungsdaten an. Wie steht Ihre Arbeitsgruppe zu diesem Themenfeld? (Mehrfachauswahl möglich)”

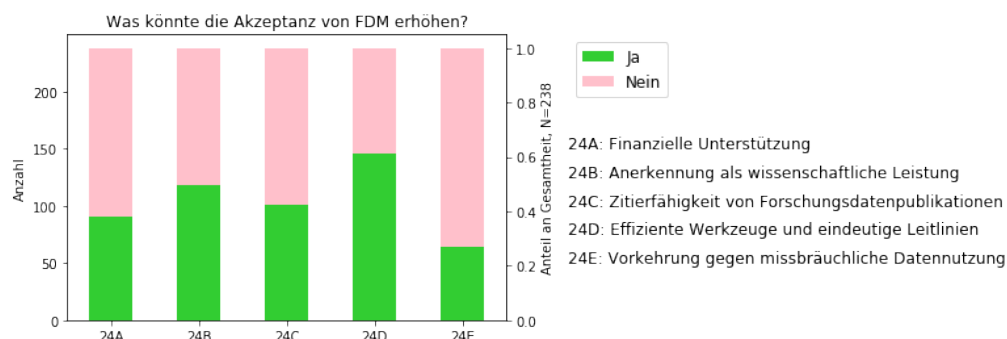


Abbildung 21: Antworten auf Frage 24: “Was könnte die Akzeptanz von Forschungsdatenmanagement erhöhen? (Mehrfachauswahl möglich)”

textfeld zur Verfügung.

Alle Teilnehmenden, die diese Frage beantworteten, entschieden sich bei allen fünf Optionen zwischen “Ja” und “Nein”.

“Effiziente Werkzeuge und eindeutige Leitlinien” wünscht sich mit 61 % der größte Anteil der Befragten. Es folgen die “Anerkennung als wissenschaftliche Leistung” mit 50 % und die “Zitierfähigkeit von Forschungsdatenpublikationen” mit 42 %. “Finanzielle Unterstützung” wird von 38 % als ausschlaggebend betrachtet. “Vorkehrungen gegen missbräuchliche Datennutzung” sehen nur 27 % der Befragten als wichtig um die Akzeptanz von FDM-Lösungen zu erhöhen.

### 3.21.1 Freitext-Antworten

Wir zeichneten 30 Freitext-Antworten auf, d.h. von 13 % derer, die die Frage beantworteten. Die Gruppierung der Antworten in verwandte Aussagen ergibt, dass am häufigsten ein konkreter Mehrwert des FDM für den Forschungsworkflow gewünscht wird. Die folgende Aussage steht prototypisch für diese Klasse, die sich mit der vorgegebenen Option “Effiziente Werkzeuge und eindeutige Leitlinien” überschneidet:

*“FDM muss einen echten Mehrwert (schnellerer Zugriff / bessere Filterung auf Daten) bieten, ohne den experimentellen Workflow zu verlangsamen.”*

Weiter lassen sich, in absteigender Häufigkeit, mehrere Antworten zu den folgenden Bedarfs-Kategorien zusammenfassen: Standards und Policies, technischer und personeller Support sowie Wahlfreiheit und

*“Die Freiheit, FDM selbst zu gestalten - keine zentralen Vorgaben!”*

Letzterer Wunsch deckt sich gut mit dem *bottom-up*-Design der NFDI, wie er

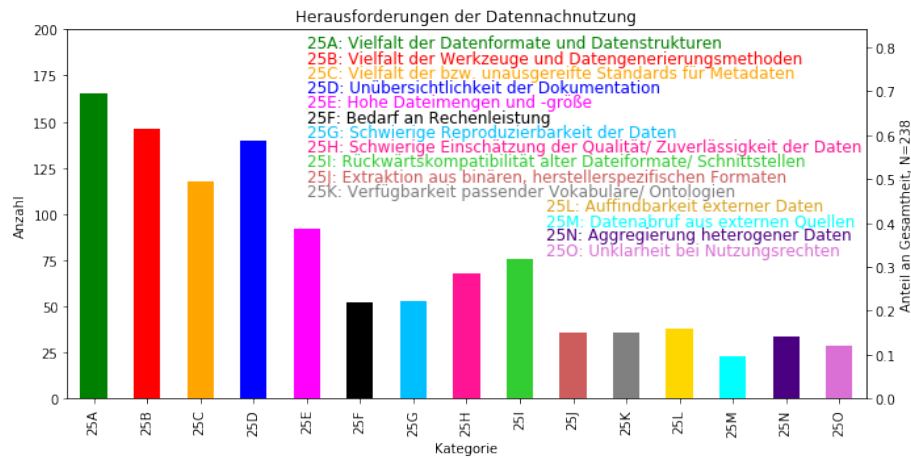


Abbildung 22: Antworten auf Frage 25: “Welche der folgenden Aspekte werden in Ihrem Fachgebiet als Herausforderung / Problem für den Datenaustausch, die Datennachnutzung empfunden? (Mehrfachauswahl möglich)”

für die NFDI gefordert und umgesetzt ist.

### 3.22 Frage 25: Herausforderungen der Datennachnutzung

Die Frage nach Herausforderungen und Problemen der Datennachnutzung zielt auf die Prioritäten einer künftigen Forschungsdateninfrastruktur.

Die “Vielfalt der Datenformate und Datenstrukturen (komplex strukturierte Datensätze)” (69 % der Antworten), die “Vielfalt der Werkzeuge und Datengenerierungsmethoden” (61 %), “Unübersichtlichkeit der Dokumentation” (59 %) sowie “Vielfalt der bzw. unausgereifte Standards für Metadaten” (50 %) werden als häufigste Erschwernisse der Datennachnutzung identifiziert. Desweiteren werden “Hohe Dateimengen und -größe” (39 %), “Rückwärtskompatibilität alter Dateiformate / Schnittstellen mit aktuell verwendeten Werkzeugen (Hardware oder Software)” (32 %) sowie “Schwierige Einschätzung der Qualität / Zuverlässigkeit der Daten” (29 %) als wichtige Problemfelder genannt. Jede der fünfzehn genannten potentiellen Herausforderungen wird von mindestens jeder zehnten antwortenden Arbeitsgruppe als relevant für die Datennachnutzung auf ihrem Forschungsgebiet angesehen.

In einem Freitextfeld konnten weitere Gründe genannt werden. Acht Teilnehmende nutzten diese Möglichkeit. Die Antworten verweisen u.a. auf fehlende Community-Standards, was bei einem bestimmten Versuch zu annotieren sei und dass es oft *a priori* unklar ist, welche Daten langfristig wertvoll sein werden. Die Abwägung von Nutzen und Kosten der Datenannotation steht dabei im Vordergrund. Dabei gehen die Antworten aber vor mehr oder minder manuellen *status quo* aus. In der NFDI geplante Entwicklungen zielen darauf hin, durch

stärkere Automatisierung dieses “Gleichgewicht” hin zu einfacherer, häufigerer und besserer Metadatenauszeichnung zu verschieben.

### 3.23 Frage 26: Weitere Anregungen

Zum Abschluß der Umfrage wurde den Teilnehmenden die Möglichkeit gegeben, in einem Freitextfeld weitere Anregungen, Kommentare und Ideen zu hinterlassen. Naturgemäß sind die Antworten sehr vielfältig und breitgefächert. Bei 35 Antworten ergaben sich bei thematischer Gruppierung analog zu Frage 22 (Abschnitt 3.19) allein 20 verschiedene Kategorien.

Viele Kommentare geben organisatorische und wissenschaftspolitische Aspekte wieder, auf welche in unserer Umfrage nicht ausführlicher eingegangen werden konnte. Die folgenden beiden Auszüge spiegeln den Tenor der Aussagen wider:

*In meinem Forschungsgebiet arbeiten wir mit extrem individualisierten experimentellen Aufbauten. Entsprechend sind die Rohdaten individualisiert. Für neue Gruppenmitglieder bedeutet das in der Regel eine mehrmonatige Einarbeitungsphase bis die Apparatur genutzt werden kann. [...]*

*Ich begrüße die Initiative sehr und sehe große Vorteile. [...] Wir haben keine Finanzierung für eine (hochqualifizierte) Personalkapazität oder gar mehr, um die Aufarbeitung von Datenstrukturen und die Vorbereitung der sehr detailreichen Analyse, Systematisierung und darauf folgender Harmonisierungsprozesse durchzuführen. [...]*

Häufig genannt werden Zeit und Effizienzabwägungen, Kommentare zur Akzeptanz strukturierten Forschungsdatenmanagements, zum Aufbau der NFDI, zu den zu setzenden Standards sowie zur Notwendigkeit, Mitarbeitende für FDM-Aufgaben zu qualifizieren (Data Literacy) – insb. vor dem Hintergrund befristeter Arbeitsverträge. Auch skeptische bis ablehnende Meinungen zur NFDI-Initiative sind mit fünf Nennungen relativ häufig. Dabei zeichnen sich einige dieser Kommentare durch reflektierte und konstruktive Kritik aus und verweisen auf die hohen (Mehr-)Anforderungen besonders an kleine Arbeitsgruppen, die mit hochgradig heterogenen und variablen Forschungsgegenständen und Experimenten arbeiten.

## 4 Vergleich mit anderen NFDI-Umfragen

In benachbarten Wissenschaften wurden bereits ähnliche FDM-Umfragen zur Vorbereitung des NFDI-Prozesses durchgeführt. Der Vergleich unserer Ergebnisse mit Erhebungen von NFDI-Konsortien in der Chemie [2] und den Ingenieurwissenschaften<sup>2,3</sup> durchgeführten Umfragen kann Aufschluß über Gemein-

<sup>2</sup><https://doi.org/10.25534/tudatalib-104>, abgerufen am 2021-0X-YZ

<sup>3</sup>[https://nfdi-matwerk.de/wp-content/uploads/2020/08/2020-08-18\\_NFDI-MatWerk\\_Digitalisierungsumfrage\\_Publikation.pdf](https://nfdi-matwerk.de/wp-content/uploads/2020/08/2020-08-18_NFDI-MatWerk_Digitalisierungsumfrage_Publikation.pdf), abgerufen am 2021-0X-YZ



samkeiten zwischen den Fächern wie auch Besonderheiten der Physik geben.

## 4.1 Vergleich zur Umfrage von NFDI4Chem

Das Konsortium NFDI4Chem führte seine Umfrage zum FDM im Fach Chemie im Rahmen der ersten NFDI-Antragsrunde im zweiten Halbjahr 2019 durch. Die Fragestellung zielte vor allem auf den gegenwärtigen Umgang mit Forschungsdaten in der Chemie, um daraus Anforderungen für die NFDI abzuleiten. Bezüglich Datentypen und -formaten machen einige Standardverfahren einen wesentlichen Teil der anfallenden Daten aus. Auch wenn innerhalb z.B. der meistgenanntesten Kategorie, "spektroskopische Daten", verschiedene Methoden und Datentypen versammelt sein können, scheint unsere Umfrage zu belegen, daß die Physik eine (noch) heterogenere Datenlandschaft aufweist.

Eine knappe Mehrheit von 53 % der von NFDI4Chem Befragten verneint die Frage, ob die Datenaufnahme von Anfang an elektronisch sei. Typischerweise geschehen Versuchplanung und Dokumentation analog, während die Analyse bei 60 % "nahtlos digital" geschieht. Dies deckt sich mit der häufig händischen Dokumentation in der Physik (siehe Abschnitt 3.13).

Bei der Frage nach Metadaten standen in der NFDI4Chem-Umfrage andere Kategorien als bei uns zur Auswahl. Die als die wichtigsten Metadaten wahrgenommenen Angaben ähneln sich jedoch in Chemie und Physik und umfassen z.B. Zeitstempel, die Probenbeschreibung und die Zuordnung zu einer Person.

Ebenso wie unsere Studie konstatiert die NFDI4Chem-Umfrage einen Mangel an Leitlinien für Datenformate und Datenorganisation sowie an Metadatenstandards. Entsprechend werden sich die NFDI-Konsortien vor ähnlichen Herausforderungen sehen, welche als *Querschnittsthemen* von Konsortien gemeinschaftlich bearbeitet werden sollen.

## 4.2 Vergleich zur Umfrage von NFDI4Ing

Die Community-Umfrage des Konsortiums NFDI4Ing vom Sommer 2019 diente bei der Erstellung unserer Umfrage als Vorlage. Daher sind Fragen teils identisch formuliert und somit besonders gut vergleichbar. Mehrere Fragen der NFDI4Ing-Umfrage erlaubten eine Auswahl nach dem Grad der Zustimmung. Zum Zwecke der Vergleichbarkeit interpretieren wir hier ein "trifft weniger zu" als eine verneinende, ein "trifft größtenteils zu" als eine bejahende Antwort.

Wie zu erwarten, nehmen in den ingenieurwissenschaftlichen Communities die angewandte Forschung und die Technologieentwicklung wichtigere Rollen ein als in der Physik, in der in Mittel mehr Grundlagenforschung betrieben wird. Was den Stand des Umgangs mit Forschungsdaten angeht, gibt es deutliche Parallelen: Physiker\*innen und Ingenieur\*innen produzieren in sehr ähnlichem Maße Text-, Bild-, und Quellcodeformate als Forschungsdaten, wobei in der Physik das durchschnittliche jährliche Datenvolumen höher ist. Die antwortenden Physiker\*innen benutzen ähnlich viel selbstgeschriebene und proprietäre Software, aber jeweils weniger Spreadsheets, Skripte und Open Source-Software als die Ingenieur\*innen. Dies könnte durch den größeren Anteil der Theorie in der Physik

zu erklären sein.

Die von uns befragten Gruppen benutzen etwas häufiger standardisierte Metadatenschemata als die von NFDI4Ing Befragten; dort ist der Anteil ohne Metadatenerfassung etwas höher. Sowohl bei den Befragten der NFDI4Phys- (27 %) als auch der NFDI4Ing-Umfrage (26 %) hat bereits ein gutes Viertel ein Repository zur Veröffentlichung von Forschungsergebnissen benutzt. Auf einzelne Projekte und Anwendungen “maßgeschneiderte” Hardware und Software wird in beiden Communities als Hindernis der (internen oder externen) Datennachnutzung erkannt.

Die in beiden Umfragen vergleichbaren Fragen beleuchten eher die Ähnlichkeiten zwischen Ingenieurwissenschaften und Physik. Während NFDI4Ing weitere Fragen nach einzelnen fachspezifischen Herausforderungen stellte (z.B. Probleme der Konnektivität bei Feldversuchen betreffend), entschied NFDI4Phys sich auch aufgrund der zum Zeitpunkt der Umfrage noch weniger klar definierten Community dafür, ein sehr breites Spektrum möglicher Probleme zu erfragen (siehe Abschnitt 3.22).

### 4.3 Vergleich zur Umfrage von NFDI-MatWerk

Die Materialwissenschaft und Werkstofftechnik siedelt sich fachlich zwischen Ingenieurwissenschaften und Physik an. Methodisch liegt der Schwerpunkt bei der experimentellen Datenerzeugung. Im Juli 2020 befragte das Konsortium NFDI-MatWerk seine Community in einer Online-Umfrage. Bezüglich der Bedarfe bei der Digitalisierung konnte für 18 Auswahlmöglichkeiten Zustimmung, eine neutrale Position oder Ablehnung angegeben werden. Die meisten dieser Bedarfe werden von deutlichen Mehrheiten geteilt. Obwohl aufgrund verschiedener Fragestellungen ein quantitativer Vergleich zu unserer Umfrage schwierig ist, sehen wir qualitativ eine ähnliches Meinungsbild. Übereinstimmend mit unserer Umfrage werden die höchsten Zustimmungswerte von jeweils  $> 80\%$  im Bereich der Datenerzeugung (Kontextinformationen, Einheitensysteme, vereinheitlichte Rohdaten sowie Software-Schnittstellen) erreicht. Aus der großen Zustimmung zu den FAIR-Prinzipien und den abgefragten Bedarfen folgert NFDI-MatWerk, dass ihre Community sich “dringend Lösungen für die digitale Zusammenarbeit” wünscht.

## Literatur

- [1] Düchs, Georg and Mecke, Klaus. Vielfalt statt Einfalt – Statistiken zum Physikstudium an den Universitäten in Deutschland 2019. *Physik Journal*, 18(8/9):26–, 2019. URL <https://www.pro-physik.de/physik-journal/august-september-2019>.
- [2] S. Herres-Pawlis, J. C. Liermann, and O. Koepler. Research data in chemistry – results of the first nfdi4chem community survey. *Z. Anorg. Allg. Chem.*, 646(21):1748–1757, 2020. URL <https://doi.org/10.1002/zaac.202000339>.