

Forschungsdaten in der Chemie

Masterarbeit

Studiengang Bibliotheks- und Informationswissenschaften (MALIS)

Fakultät für Informations- und Kommunikationswissenschaften

Technischen Hochschule Köln

Vorgelegt von

Daniela Adele Hausen

Am 16.09.2019

Gutachterin: Frau Dr. Ania López

Zweitgutachter: Herr Prof. Dr. Achim Oßwald

Picard: „Ist es eine Lebensform, Data?“

Data: „Zu wenige Informationen für eine Antwort, Captain.“

Picard: „Und theoretisch?“

Data: „Hm, es wäre möglich.“

Zitat aus Star Trek – The next generation (Staffel 1, Folge 2)

Inhaltsverzeichnis

Inhaltsverzeichnis.....	3
Abbildungsverzeichnis.....	5
Tabellenverzeichnis.....	7
Abkürzungsverzeichnis.....	8
0. Abstract.....	10
1. Einleitung	12
2. Forschungsdaten und Forschungsdatenmanagement in der Chemie.....	15
2.1 NFDI4Chem.....	17
3. Umfragen zu Forschungsdaten.....	18
4. Definition der Forschungsziele	19
5. Methodik.....	20
5.1 Zielgruppenbestimmung	21
5.1.1 Einladung der Teilnehmenden und Rücklaufquote	23
5.2 Spezifizierung des Fragebogens	26
5.2.1 Einleitung oder Einführungstext	26
5.2.2 Fragen und Antwortvorgaben sowie offene Texteingaben.....	26
5.2.3 Konkrete und verständliche Hinweise	27
5.3 Durchführung und Ergebnisse des Pretests	29
6 Umsetzung.....	29
6.1 SoSciSurvey.....	30
7. Auswertung der erhobenen Daten.....	30
7.1 Position der Teilnehmenden	31
7.2 Ist-Zustand	33

7.3	Datenarten	35
7.4	Auffindbarkeit.....	40
7.5	Verständlichkeit der Daten und Laborbücher für den Datenzugriff	47
7.6	Strukturierung der Daten und Vorgaben	54
7.7	Datenweitergabe und Datenpublikation.....	58
7.8	Danksagung und Anmerkungen	67
8.	Zusammenfassung und Ausblick.....	67
9.	Literaturverzeichnis.....	70
10.	Danksagung	77
11.	Anhang	78
11.1	Einladungsanschreiben.....	78
11.2	Erinnerungsanschreiben.....	79
11.3	Fragebogen.....	80
12	Eidesstaatliche Erklärung	86

Abbildungsverzeichnis

Abbildung 1: Forschungsdatenlebenszyklus (eigene Darstellung)	13
Abbildung 2: FAIR-Prinzipien by Sangya Pundir	13
Abbildung 3: Rücklaufstatistik über den Zeitraum vom 06.05. bis 14.06.2019 (Quelle: www.socsisurvey.de)	24
Abbildung 4: Screenshot der Datensätze und Abbrüche (Quelle: www.socsisurvey.de)	25
Abbildung 5: Auswertung der Frage - Welche Position haben Sie an Ihrem Institut inne?	32
Abbildung 6: Auswertung der Frage - Kennen Sie die FAIR-Prinzipien? Ist Ihnen die folgende Grafik vertraut?	33
Abbildung 7: Auswertung der Frage - „Welche Daten erheben Sie bei Ihrer Forschung?“; Anzahl gewählter Antwortoptionen & relative Häufigkeit der Teilnehmenden	36
Abbildung 8: Auswertung der Frage – Welche Daten erheben sie bei Ihrer Forschung?	36
Abbildung 10: Auswertung der Frage - „Welche Versionen ihrer Daten speichern Sie am Projektende ab?“; Anzahl gewählter Antwortoptionen & relative Häufigkeit der Teilnehmenden	38
Abbildung 11: Auswertung der Frage DA02- verschiedene Datenarten zur Aufbewahrung	39
Abbildung 12: Auswertung der Frage - „Mit welchen Metadaten beschreiben Sie Ihre Daten?“; Anzahl gegebener Metadaten & relative Häufigkeit der Teilnehmenden	41
Abbildung 13: EVA-Prinzip	43
Abbildung 14: Auswertung der Frage – „Könnte eine Wissenschaftlerin/ ein Wissenschaftler innerhalb oder außerhalb ihres Instituts mit den bisher von Ihnen mitgelieferten Metadaten, Ihre Daten verstehen?“; Antwortmöglichkeiten & relative Häufigkeit der Teilnehmenden	48
Abbildung 15: Auswertung zur Frage - „Was für eine Art Laborbuch (Papier oder elektronisches Laborbuch) verwenden Sie?	50
Abbildung 16: Gründe zur Verwendung eine Laborbuchs in Papierformat	51
Abbildung 17: Gründe zur Verwendung eines elektronischen Laborbuchs	53
Abbildung 18: Gründe zur Verwendung keines Laborbuchs	53

Abbildung 19: Auswertung der Frage – „Wie strukturieren bzw. benennen Sie Ihre Daten?“; Anzahl gewählter Antwortoptionen & relativen Häufigkeit der Teilnehmenden	54
Abbildung 20: Auswertung der Frage I001 – Wie strukturieren bzw. benennen Sie ihre Daten?	55
Abbildung 21: Auswertung der Frage - „Strukturieren und/ oder beschreiben Sie Ihre Daten mit Metadaten?“, Anzahl gewählter Antwortoptionen & relativen Häufigkeit der Teilnehmenden	56
Abbildung 22: Auswertung der Frage - „Strukturieren und/ oder beschreiben Sie Ihre Daten mit Metadaten?“	57
Abbildung 23: Auswertung der Frage – „Wie teilen Sie Daten innerhalb Ihres Instituts?“; Anzahl gewählter Antwortoptionen & relative Häufigkeit	59
Abbildung 24: Auswertung der Frage – „Wie teilen Sie Daten innerhalb Ihres Instituts?“	60
Abbildung 26: Auswertung der Frage – „Wie teilen Sie Daten außerhalb Ihres Instituts?“	61
Abbildung 25: Auswertung der Frage – „Wie teilen Sie Daten außerhalb Ihres Instituts?“; Anzahl gewählter Antwortoptionen & relative Häufigkeit der Teilnehmenden	61
Abbildung 27: Auswertung der Frage – „Haben Sie bereits Daten veröffentlicht?“; Anzahl gewählter Antwortoptionen & relative Häufigkeit der Teilnehmenden	62
Abbildung 28: Auswertung der Frage – „Haben Sie bereits Daten veröffentlicht?“	63
Abbildung 29: Veröffentlichung eines ¹ H-NMR-Spektrums von Yichao Yan et al., „Supporting Information for: Mechanism-based design of a high-potential catholyte enables a 3.2 V all-organic nonaqueous redox flow battery,“ Journal of the American Chemical Society, 2019.	64
Abbildung 30: Auswertung der Frage – „Haben Sie selbst schonmal Daten aus anderen Quellen nachgenutzt?“; Anzahl gewählter Antwortoptionen & relative Häufigkeit der Teilnehmenden	65
Abbildung 31: Auswertung der Frage - „Haben Sie selbst schonmal Daten aus anderen Quellen nachgenutzt?“	66
Abbildung 32: durchschnittlich gegebene Antworten	69
Abbildung 32: Fragebogen in SoSciSurvey - Fragen 1 und 2.....	80

Abbildung 33: Fragebogen in SoSciSurvey – Frage 3	80
Abbildung 34: Fragebogen in SoSciSurvey - Frage 4.....	81
Abbildung 35: Fragebogen in SoSciSurvey - Frage 5.....	81
Abbildung 36: Fragebogen in SoSciSurvey - Fragen 6 und 7.....	82
Abbildung 37: Fragebogen in SoSciSurvey - Frage 8.....	82
Abbildung 38: Fragebogen in SoSciSurvey - Frage 9.....	83
Abbildung 39: Fragebogen in SoSciSurvey - Frage10.....	83
Abbildung 40: Fragebogen in SoSciSurvey - Fragen 11 und 12.....	84
Abbildung 41: Fragebogen in SoSciSurvey - Frage 13 und Anmerkungen.....	84
Abbildung 42: Fragebogen in SoSciSurvey - Danksagung	85

Tabellenverzeichnis

Tabelle 1: Potentiell Teilnehmenden der Befragung	24
Tabelle 2: Korrelation Position zum Wissen um die FAIR-Prinzipien/ -Grafik	34
Tabelle 3: Bewertung der genannten Metadaten von 0 bis 9 (0 = unbedeutend und 9 = sehr wichtig).....	45
Tabelle 4: Korrelation: Verständlichkeit der Daten zu Position der Teilnehmenden	49
Tabelle 5: am meisten genannte Metadaten bei der offenen Texteingabe	56

Abkürzungsverzeichnis

AC	Anorganische Chemie
DCC	Digital Curation Centre
DFG	Deutsche Forschungsgemeinschaft
DSGVO	Datenschutzgrundverordnung
EUDAT	European Data Project
ELN	Elektronisches Laborbuch
EOSC	European Open Science Cloud
Etc.	et cetera
FDM	Forschungsdatenmanagement
Ggfs.	Gegebenenfalls
GPC	Gelpermeationschromatographie
HHU	Heinrich-Heine-Universität Düsseldorf
IR-Spektroskopie	Infrarot-Spektroskopie
LIMS	Laborinformationssystem
NFDI	Nationale Forschungsdateninfrastruktur
NFDI4Chem	Fachkonsortium Chemie für die Nationale Forschungsdateninfrastruktur
NMR	Kernspinresonanz
NRW	Nordrhein-Westfalen
OC	Organische Chemie
PC	Physikalische Chemie

Forschungsdaten in der Chemie

RDA	Research Data Alliance
RfII	Rat für Informationsinfrastrukturen
RWTH	Rheinisch-Westfälische Technische Hochschule
z.B.	zum Beispiel

0. Abstract

Ziel der vorliegenden Arbeit ist es, den aktuellen Stand zu einzelnen Aspekten des Forschungsdatenmanagements im Fachbereich Chemie an Nordrhein-Westfälischen Universitäten zu untersuchen. Als Erhebungsinstrument wurde eine quantitative Online-Umfrage eingesetzt. Aufgrund der unbekanntem Grundgesamtheit lassen sich keine validierten Aussagen treffen, aber es zeichnen sich teilweise sehr eindeutige Trends ab.

Fokussiert wurde in der Umfrage auf die Bereiche Auffinden, Strukturieren, Teilen und Veröffentlichen von Daten. Hierbei wurde unter anderem untersucht, wie weit die elektronischen Laborbücher schon verbreitet sind. Tangiert wurden ebenfalls die Nachnutzbarkeit und Verständlichkeit von Daten aus Sicht der Forschenden. Bei der Auffindbarkeit und Strukturierung wurden die Forschenden gebeten, ihre verwendeten Metadaten zu nennen. Diese Metadaten wurden nach dem EVA-Prinzip geclustert und ausgewertet, dabei zeigte sich, dass das wichtigste Metadatum der Teilnehmenden das Datum ist. Als weiteres signifikantes Ergebnis stellte sich raus, dass die meisten Forschenden die FAIR-Prinzipien nicht kennen. Insgesamt wurde deutlich, dass es für den Fachbereich Chemie noch ein langer Weg ist zu FAIRem Forschungsdatenmanagement.

Die ausgewerteten Daten sind auf Zenodo unter der DOI [10.5281/zenodo.3415059](https://doi.org/10.5281/zenodo.3415059) unter einer CC-BY-NC 4.0 Lizenz veröffentlicht.

Schlagwörter: Forschungsdatenmanagement, Chemie, Metadaten, NFDI4Chem, FAIR, EVA-Prinzip, Online-Umfrage, Laborbuch, ELN

The aim of the present study is to examine the current status of several aspects of research data management in the departments of chemistry at North Rhine-Westphalian universities. A quantitative online survey has been conducted. Due to the unknown population of the participants, no validated statements can be made, but it is possible to derive very clear trends.

The survey focused on finding, structuring, sharing and publishing research data. Among other things, it was examined how common the electronic lab notebooks really are. The reusability and comprehensibility of data from the perspective of the researchers were also tangented. In the questions of finding and structuring, researchers were asked to name their common used metadata. These metadata were clustered and evaluated according to the EVA principle, showing that the most important metadata of the participants is the date. Another significant result was that most researchers did not know the FAIR principles. Overall, it became clear that chemistry has still a long way to go for FAIR research data management.

The dataset is available on Zenodo: DOI [10.5281/zenodo.3415059](https://doi.org/10.5281/zenodo.3415059) under CC-BY-NC 4.0.

Keywords: Research Data Management, chemistry, metadata, NFDI4Chem, FAIR, online survey, lab journal, ELN

1. Einleitung

Das Thema Forschungsdatenmanagement (FDM) hat in den vergangenen Jahren an deutschen Hochschulen stark an Bedeutung gewonnen. Die stetig wachsenden Datenmengen, die in der Forschung produziert und verarbeitet werden, und das zunehmende Bewusstsein für den Wert von Daten als Grundlage für Erkenntnisgewinn haben die Anforderungen an Forschende gesteigert, ihre Daten professionell zu managen, langfristig zu sichern und zunehmend auch öffentlich zugänglich zu machen¹. Doch was sind genau Forschungsdaten? Eine eindeutige Definition des Begriffs gibt es bisweilen nicht, sondern eine Vielzahl an Definitionen². Die Deutsche Forschungsgemeinschaft (DFG) hat 2015 Forschungsdaten wie folgt beschrieben:

„Zu Forschungsdaten zählen u.a. Messdaten, Laborwerte, audiovisuelle Informationen, Texte, Surveydaten, Objekte aus Sammlungen oder Proben, die in der wissenschaftlichen Arbeit entstehen, entwickelt oder ausgewertet werden. Methodische Testverfahren, wie Fragebögen, Software und Simulationen können ebenfalls zentrale Ergebnisse wissenschaftlicher Forschung darstellen und sollten daher ebenfalls unter den Begriff Forschungsdaten gefasst werden.“ (DFG, 2015)³

Diese Aufzählung verdeutlicht bereits sehr gut die heterogene Ansicht was Forschungsdaten sind. Die Zusammenstellung weiterer möglicher Definitionen auf forschungsdaten.org⁴ zu Forschungsdaten zeigt nachdrücklich, dass kein Konsens über die Definition herrscht, sondern diese aus Sicht einzelner Disziplinen und Fördermittelgeber geprägt ist⁵.

Deutlich kongruenter zu sein erscheint es beim Management von Forschungsdaten. Dieses wird als Aufgabe verstanden, strukturierte und langfristige Ansätze zum Umgang mit der exponentiell zunehmenden Menge digitaler Forschungsdaten und der damit einhergehenden Steigerung der Komplexität zu schaffen⁶. Verschiedene Modelle wie z.B. das Domänen-Modell⁷ von *Treloar* und *Harboe-Rae* oder der

¹ Winkler-Nees 2011.

² forschungsdaten.org o.J.

³ Allianz der deutschen Wissenschaftsorganisationen 2010.

⁴ forschungsdaten.org o.J.

⁵ University of Leicester o.J.

⁶ Hochschulrektorenkonferenz 2014.

⁷ Treloar und Harboe-Rae 2008.

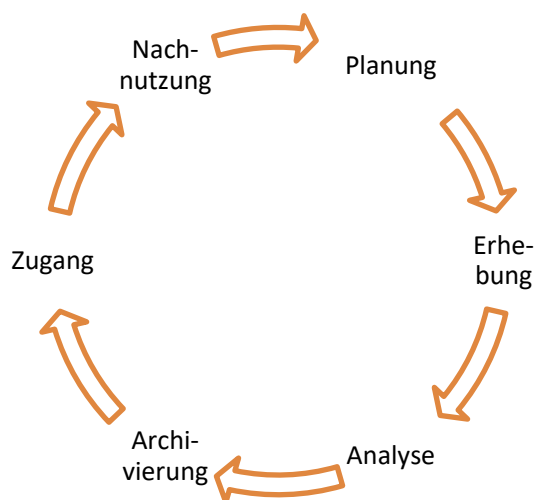


Abbildung 1: Forschungsdatenlebenszyklus (eigene Darstellung)

Forschungsdatenlebenszyklus des *Digital Curation Centers* (DCC)⁸ beschreiben und visualisieren die Aufgaben, Prozesse und das Ziel des FDMs. In Abbildung 1 ist ein vereinfachtes Modell des Forschungsdatenlebenszyklus⁹ gezeigt, das verdeutlicht, dass FDM bereits vor der Datenerhebung und Analyse mit der Planung beginnt.

Bei den ersten drei Schritten müssen bereits Überlegungen und Maßnahmen¹⁰ einfließen, um eine anschließende Archivierung, Zugänglichkeit sowie Nachnutzung der Daten zu garantieren. Es kristallisiert sich die (Nach-) Nutzung – in öffentlicher oder geschlossener Umgebung – der Daten als Ziel des FDMs heraus¹¹. Um dieses Ziel zu erreichen, stellen die FAIR-Prinzipien¹² (vgl. Abbildung 2) Rahmenbedingungen dar. Das Akronym FAIR steht für Findable, Accessible, Interoperable und Reusable.

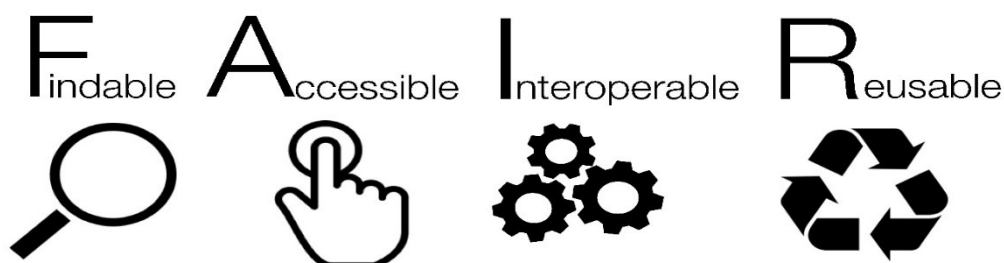


Abbildung 2: FAIR-Prinzipien by Sangya Pundir

FAIR bezieht sich nicht allein auf die Daten, sondern ebenfalls auf die Tools, die Workflows und Algorithmen, die mit den Daten zusammenhängen. Das Ziel der FAIR-Prinzipien ist eine menschen- und maschinenlesbare Aufbereitung der Daten, dass diese für weitere Forschung nachgenutzt werden können¹³.

⁸ Digital Curation Center 2004-2019.

⁹ Eigene Darstellung

¹⁰ Hausen und Trautwein-Bruns 2019.

¹¹ Ludwig und Enke 2013.

¹² Wilkinson et al. 2016.

¹³ Wilkinson et al. 2016.

Um dieses Ziel der Nutzung von Forschungsdaten zu erreichen, sind neue Infrastrukturen erforderlich, womit, der Definition des *Rates für Informationsinfrastrukturen* (RfII) folgend, nicht nur leistungsfähige IT-Systeme zur Verarbeitung von Daten, sondern auch Prozesse, Kompetenzen und eine entsprechende Datenkultur gemeint sind¹⁴. Auf dem Weg zu diesem Ziel haben viele Universitäten Unterstützungsangebote für das FDM entwickelt, die neben IT-Lösungen auch Beratungs-, Weiterbildungs- und Awareness-Maßnahmen umfassen. Die Bemühungen der Hochschulen werden in Deutschland in vielen Bundesländern durch regionale Projekte und Strukturen wie der *Landesinitiative NFDI der Digitalen Hochschule NRW*¹⁵ oder *Hessische Forschungsdateninfrastrukturen* (HeFDI)¹⁶ unterstützt, koordiniert und teilweise auch getrieben. Auf europäischer Ebene wurden mit der *European Open Science Cloud* (EOSC)¹⁷, dem *European Data Project* (EUDAT)-Projekt¹⁸ und dem *Open Research Data Pilot* für EU-geförderte Forschungsprojekte ebenfalls großdimensionierte Initiativen gestartet, die dem Ziel verpflichtet sind, den Umgang mit Daten in der Forschung grundlegend zu verändern. Gleichzeitig wurde innerhalb der wissenschaftlichen Fachgemeinschaften ein Reflexions- und Verständigungsprozess über die Frage, was gutes Datenmanagement in der jeweiligen Disziplin bedeutet, begonnen.

Um die lokale, regionale und europäische Ebene adäquat zu verknüpfen, soll auf Bundesebene in Deutschland eine Nationale Forschungsdateninfrastruktur (NFDI), wie sie durch den RfII zur Diskussion gestellt wurde, und gemeinsam mit der *Gemeinsamen Wissenschaftskonferenz des Bundes und der Länder* (GWK) vorangetrieben wird, entstehen. Vorhandene Akteure, bestehende Informationseinrichtungen, Infrastrukturen, Angebote, usw. werden aufgegriffen, miteinander verknüpft und ausgebaut sowie Lücken geschlossen werden. Betont wird, dass es sich um keine weitere „Säule“ im Wissenschaftssystem handeln wird¹⁹, sondern ein disziplin- und community-übergreifendes Netzwerk entstehen wird. Die Konsortien, die im Rahmen der NFDI gebildet werden, sollen von der jeweiligen wissenschaftlichen Community/

¹⁴ Rat für Informationsinfrastrukturen 2016.

¹⁵ Universität Duisburg-Essen o.J.

¹⁶ Philipps-Universität Marburg o.J.

¹⁷ European Open Science Cloud 2018.

¹⁸ EUDAT - European Data Project o.J.

¹⁹ Rat für Informationsinfrastrukturen 2018.

Fachgemeinschaft mit aus ihrer Sicht geeigneten Infrastrukturpartnern definiert werden²⁰. Wie viele andere Disziplinen hat auch das Fachgebiet Chemie die Chancen, die ein NFDI bietet, erkannt und formiert sich zu einem *Fachkonsortium Chemie für die Nationale Forschungsdateninfrastruktur* (NFDI4Chem). Im Juli 2019 hat die Initiative, wie 21 weitere Initiativen auch, eine verbindliche Absichtserklärung²¹ für die Antragstellung im Jahr 2019 zum Aufbau der NFDI eingereicht²² (vgl. Kapitel 2.1).

2. Forschungsdaten und Forschungsdatenmanagement in der Chemie

Forschungsdaten in der Chemie weisen aufgrund der vielen Teildisziplinen und der damit verbundenen Vielzahl an unterschiedlichen Erhebungs- und Analysemethoden eine hohe Heterogenität auf²³. Diese sowie die Komplexität der Anforderungen an die Daten wird durch mit anderen Disziplinen überlappende Teildisziplinen wie z.B. der physikalischen Chemie (PC) oder der Biochemie gesteigert. Forschungsdatenarten in der Chemie können z.B. über das Merkmal Methode allgemein auf experimentelle, analytische und Simulationsdaten heruntergebrochen werden. Diese liegen häufig in proprietären Formaten und ohne oder nur mit beschränkten, beschreibenden Metadaten vor²⁴.

Wie auch in anderen Disziplinen werden in der Chemie Forschungsdaten generiert, die potentiell zur Nachnutzung geeignet sind, wenn sie auffindbar, zugänglich und durch ausreichende Metadaten verständlich beschrieben werden. Hierzu fehlen allerdings an vielen Forschungseinrichtungen etablierte Strukturen, um den Umgang mit den Daten zu homogenisieren. Ein Mangel an Standards und Kriterien zur Strukturierung, Vernetzung, Speicherung, Kuratation und Veröffentlichung der Daten ist die Folge²⁵.

Dabei ist die Kristallographie als Teildisziplin der Chemie ein prominentes und sehr positives Beispiel für die Datenpublikation. Seit über 50 Jahren werden Einkristallstrukturen auf Repositorien wie z.B. der *Cambridge Structural Database* (CSD)

²⁰ Rat für Informationsinfrastrukturen 2018.

²¹ Initiative NFDI4Chem 2019.

²² Deutsche Forschungsgemeinschaft 2019.

²³ Koepler et al. 2018.

²⁴ Koepler et al. 2018.

²⁵ Koepler et al. 2018.

des *Cambridge Crystallographic Data Centre* (CCDC) oder der *Crystallography Open Database* (COD) abgelegt. Seit der Gründung der CSD 1965 wird eine systematische Kuration der Daten verfolgt. Mit dem Bereitstellen der Einkristallstrukturen als elektronische Ressource wurde der *Crystallographic Information File* (CIF) eingeführt, der sich mittlerweile als Datenstandard etabliert hat²⁶.

Weitere 24 Standards sind unter *fairsharing.org*²⁷ gelistet. Unter anderem der bekannte Spektrenstandard JCAMP-DX²⁸, der ein aus den 1980er Jahren offenes, maschinenlesbares Austauschformat darstellt. Die NMRReDATA²⁹ Initiative stellte 2018 als Alternative einen offenen von menschen- und maschinenlesbaren Datenstandard vor, der im internationalen Kontext erarbeitet wurde. Dieser Datenstandard verlinkt zusätzlich NMR-Parametern, wie z.B. die chemische Verschiebung oder 2D-Korrelation mit Strukturdaten und wird in einem erweiterten Molfile-Format gespeichert. Die gemessenen spektralen Daten und der NMRReDATA Datensatz werden für die Speicherung in einem Repositorium zu einem Kernspinresonanz (NMR) – Datensatz fusioniert³⁰. Insgesamt finden sich auf *fairsharing.org* 73 Einträge unter dem Begriff »chemistry«, die nach standard type, record type, domains, subjects, taxonomies, countries und organisations facettiert sind³¹. Bei der Entwicklung vieler Standards wie z.B. dem »Minimum Information About a Simulation Experiment (MIASE)³²« zeigt sich, dass es eine internationale kooperative Arbeit war. International wird das Thema durch das *IUPAC Committee on Publications and Cheminformatics Data Standards* (CPCDS)³³, die *Chemistry Research Data Interest Group* (CRDIG)³⁴ der *Research Data Alliance* sowie dem *Go FAIR Chemistry Implementation Network* (ChIN)³⁵ vorangetrieben. 2018 haben Vertreter der zuvor genannten Gruppen zur *Data Interest Group Chemistry* (DIGChem)³⁶ zusammengefunden, um Datenstandards und Anforderungen für Daten-Repositorien in

²⁶ Bruno und Groom 2014.

²⁷ The FAIRsharing team und Oxford e Research Centre, University of Oxford 2009-2019.

²⁸ IUPAC o. J.

²⁹ Pupier et al. 2018.

³⁰ Koepler et al. 2018.

³¹ The FAIRsharing team und Oxford e Research Centre, University of Oxford 2009-2019.

³² Waltemath et al. 2011.

³³ IUPAC 2019.

³⁴ McEwen et al. 2019.

³⁵ Coles o.J.

³⁶ IUPAC 2018.

der Chemie zu diskutieren und zu entwickeln. In Deutschland wird FDM durch die Initiative NFDI4Chem vorangebracht.

2.1 NFDI4Chem

Erste Überlegungen zur NFDI4Chem gab es bei einer Auftaktveranstaltung »Fachgespräch NFDI4Chem« im geschlossenen Kreis im April 2018, an der Forschende aus der Chemie, die Fachgesellschaft *Gesellschaft Deutscher Chemiker* (GDCh), Einrichtungen aus der Forschungsförderung und die *Technische Informationsbibliothek* (TIB) als Infrastruktureinrichtung teilgenommen haben. Resultat war u.a. das Thesenpapier Nationale Forschungsdateninfrastruktur für die Chemie (NFDI4Chem)³⁷ und das definierte Ziel zur Schaffung einer funktionierenden Basisinfrastruktur. Konkretisiert wird dieses in der verbindlichen Absichtserklärung³⁸:

„The vision of NFDI4Chem is the digitalisation of all key steps in chemical research. NFDI4Chem supports scientists in their efforts to collect, store, process, analyse, disclose and re-use research data.“ (NFDI4Chem, Binding Letter of Intent, 2019)

Dieses Ziel soll durch den Aufbau einer virtuellen Forschungsumgebung mit verknüpften Repositorien, dem Erarbeiten von Minimum Information Standards für Daten und Metadaten im internationalen Kontext sowie durch die Förderung des kulturellen und digitalen Wandels erfolgen. Weiterhin soll Awareness für FAIRes Datenmanagement und ein rechtssicherer Rahmen geschaffen werden sowie Synergien mit anderen Konsortien entstehen. Diese Teilziele/ -aufgaben sollen durch 6 Arbeitsbereiche erreicht werden³⁹.

Sprecher der Initiative ist *Prof. Dr. Christoph Steinbeck* von der *Friedrich Schiller Universität Jena*⁴⁰, der an einem Gast-Editorial in der *Angewandten Chemie*⁴¹ mitgewirkt hat. Dieses stellt die Initiative NFDI4Chem vorstellt und ruft zu einer großen Umfrage⁴² zum Forschungsdatenmanagement in der Chemie auf.

³⁷ Koepler et. al. (2018)

³⁸ Initiative NFDI4Chem 2019.

³⁹ Initiative NFDI4Chem 2019.

⁴⁰ Howes 2019.

⁴¹ Herres-Pawlis et al. 2019.

⁴² NFDI4Chem Online Survey 2019.

3. Umfragen zu Forschungsdaten

Die von der NFDI4Chem Initiative aufgesetzte Umfrage zum FDM in der Chemie geht Fragestellungen zur Datenerhebung, -beschreibung, -teilen und zur Datenpublikation nach. Außerdem werden NFDI relevante Fragen am Ende gestellt. Die Umfrage ist nach der in dieser Arbeit vorgestellten Befragung und der Konzeptstudie zu „*Vernetzte Primärdaten-Infrastruktur für den Wissenschaftler-Arbeitsplatz in der Chemie*“⁴³ die Dritte explizit zur Forschung im Fachbereich Chemie.

Die Schwierigkeit von Online-Umfragen ist unter anderem, valide Ergebnisse zu erzeugen, da häufig die Grundgesamtheit nicht bekannt ist. In dieser Arbeit ist auch nur eine Annäherung der Grundgesamtheit möglich⁴⁴ (vgl. Kapitel 5).

Trotz dieses häufig auftretenden Nachteils ist die Anzahl der Umfragen zum Forschungsdatenmanagement groß. Der Wiki-Eintrag „Umfrage zum Umgang mit Forschungsdaten an wissenschaftlichen Institutionen“ auf *forschungsdaten.org* spiegelt mit 17 Einträgen einen Bruchteil wider. Wird mittels einer Suchmaschine nach den Termini »Umfrage« und »Forschungsdatenmanagement« gesucht, so werden 8760 Ergebnisse angezeigt⁴⁵. Es zeigt sich, dass annähernd jede Universität und außeruniversitäre Forschungseinrichtung in Deutschland mindestens eine Bestandsaufnahme zu den lokalen Forschungsdaten und den Umgang mit ihnen durchgeführt hat. Zusätzlich wurden Befragungen in verschiedenen Bundesländern wie z.B. Nordrhein-Westfalen^{46,47}, Baden-Württemberg⁴⁸ und Hessen⁴⁹ zum FDM umgesetzt. Bei den Umfragen waren die primären Zielgruppen jeweils die Forschenden der eigenen Hochschule, außeruniversitären Einrichtung oder des eigenen Landes oder zu einem geringeren Anteil Vertreterinnen und Vertreter aus Infrastruktureinrichtungen. In den primären Zielgruppen waren Chemikerinnen und Chemiker inkludiert, allerdings in den seltensten Fällen bei der Auswertung ausgewiesen. Die Ergebnisse werden zusammengefasst als Natur-, Lebens-, Geistes- und

⁴³ Technische Informationsbibliothek et al. 2010.

⁴⁴ Müller et al. 2014.

⁴⁵ Suche Umfrage und Forschungsdatenmanagement 2019.

⁴⁶ Brenger et al. 2019.

⁴⁷ Winter 2019.

⁴⁸ Tristram 2015.

⁴⁹ Brand et al. 2018.

Sozial- und Ingenieurwissenschaften, ... ausgegeben. Als Methoden wurden überwiegend Experteninterview und E-Mail bzw. Online-Umfragen eingesetzt.

In der Suche werden auch Treffer zu disziplinspezifischen Umfragen angezeigt wie z.B. Umfragen zu Forschungsdaten in den Afrika-⁵⁰, Orient-⁵¹ und Geistes- und Sozialwissenschaften⁵², wobei die zwei Ersten die NFDI im Einleitungstext zur Umfrage erwähnen.

4. Definition der Forschungsziele

Motiviert durch die Bewegungen zur disziplinspezifischen Ausrichtung des FDMs und der Lücke bei der Befragung zum Thema FDM in der Chemie, thematisiert diese Arbeit die Forschungsdaten und das Forschungsdatenmanagement in der Chemie.

Ziel dieser Arbeit ist, den aktuellen Stand zu einzelnen Aspekten des Forschungsdatenmanagements im Fachbereich Chemie an Nordrhein-Westfälischen Universitäten zu untersuchen.

Zur Orientierung werden folgende Leitfragen formuliert:

- (1) Wie weit sind die FAIR-Prinzipien in der Chemie verbreitet?
- (2) Sind die Forschungsdaten in der Chemie auffindbar?
- (3) Aus der Perspektive der Teilnehmenden sind ihre Daten verständlich und nachnutzbar?
- (4) Wie sieht das Verhalten der Beteiligten bzgl. des Datenteilens und –nachnutzens aus?

Da die Datenerhebung überregional an nordrhein-westfälischen Universitäten erfolgen soll, wird als Erhebungsinstrument die Online-Umfrage gewählt und mit dem Tool SoSciSurvey umgesetzt.

⁵⁰ Vereinigung für Afrikawissenschaften in Deutschland 2019.

⁵¹ Menalib o.J.

⁵² Daudrich und Anna 2018.

5. Methodik

Bei Online-Umfragen kommen die gleichen allgemeinen methodischen Vorgaben^{53,54} zum Tragen wie z.B. bei der mündlichen, postalischen oder telefonischen Befragung. Die Online-Befragung zählt zu den quantitativen Methoden⁵⁵ und war bereits im Jahr 2010 mit 38% die am weitesten verbreitete Methode aller quantitativen Befragungsformen⁵⁶. Sie weist im Vergleich zu den Offline-Befragungsformen Vorteile wie eine Orts- und Zeitunabhängigkeit der Teilnahme der Befragten, Zeiteffizienz sowie ein schneller Rücklauf der beantworteten Fragebögen auf. Das Einbinden von Multimedia-Elementen, eine automatische Filterführung, eine hohe Datenqualität und das Randomisieren der Antwortmöglichkeiten sowie das Vermeiden von Eingabefehlern stellen weitere technische Vorteile dieser Umfrageform dar^{57,58}. *Thielsch und Weltzin* listen den Zeitaufwand der Programmierung des Fragebogens, keine Kontrolle über die Durchführungsbedingungen sowie Mehrfachteilnahmen und die Asynchronität der Antworten auf Rückfragen als nachteilig⁵⁹. Des Weiteren sind die möglichen Verständnis- und Darstellungsprobleme zu erwähnen. Bei Befragungen ohne Interviewer muss der Befragte selbständig die Fragen verstehen und beantworten. Dieses kann durch eine verständliche sprachliche Formulierung sowie eine optische Gestaltung erleichtert werden^{60,61}. Beide Kriterien stellen zusammen mit einer klaren Filterführung, vordefinierten Eingabefeldern und verständlichen Instruktionen die qualitätsrelevanten Merkmale von Online-Befragungen dar. Hinzu wird das Einrichten eines Zurück-Buttons empfohlen, so dass bereits gegebene Antworten korrigiert werden können.⁶²

Neben den qualitätsrelevanten Merkmalen hat der *Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V.* (ADM) zusammen mit der *Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e.V.*, dem *Berufsverband Deutscher Markt- und*

⁵³ Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V. et al. 1999.

⁵⁴ Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V. et al. o.J.

⁵⁵ Baur und Blasius 2014.

⁵⁶ Thielsch und Weltzin 2012.

⁵⁷ Fühles-Ubach 2013a.115.

⁵⁸ Thielsch und Weltzin 2012.

⁵⁹ Thielsch und Weltzin 2012.

⁶⁰ Fühles-Ubach 2013a.

⁶¹ Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V. et al. o.J.(2001).

⁶² Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V. et al. o.J.

Sozialforscher e.V. und der *Deutschen Gesellschaft für Online-Forschung e.V.* zwei Kriterien für die Repräsentativität einer Online-Befragung formuliert. Zum einen muss sich die Umfrage auf eine „sachlich, regionale und zeitlich, klar definierte Grundgesamtheit“ beziehen und die Befragten müssen individuell angesprochen werden⁶³. Sowohl die Grundgesamtheit wie auch die individuelle Ansprache sind in dieser Arbeit annähernd erfüllt, lediglich bei 14 % aller Befragten aus der Grundgesamtheit lagen kein Name und keine Emailadresse vor (vgl. Kapitel 5.1). Weitere Überlegungen zu repräsentativen Ergebnissen werden von *Zerr* in zwei Grundbedingungen formuliert. In der ersten bezieht er sich auch auf die „interessierte Grundgesamtheit“⁶⁴ und in der zweiten führt er aus, dass die Ergebnisse auf diese Grundgesamtheit verallgemeinerbar sein müssen. Dieses ist am besten durch „valide sekundärstatistische Daten“⁶⁵ oder abgeschwächt durch die Strukturgleichheit der Stichprobe mit der Grundgesamtheit bezogen auf ein Erhebungsmerkmal möglich.⁶⁶ Da teilweise vergleichbare sekundärstatistischen Daten aus der Chemie fehlen und die Strukturgleichheit nicht vollständig gegeben ist (vgl. Kapitel 7.1), muss die durchgeführte Online-Umfrage als nicht repräsentativ eingestuft. Die Umfrage liefert Informationen zum Wissen um die FAIR-Prinzipien und dem Verhalten von Chemikerinnen und Chemikern zu ihren Daten und zum FDM, die durch teilweise sehr eindeutige Ergebnisse der Fragen zu den Erhebungsmerkmalen einen klaren Trend aufzeigen⁶⁷.

Die Umfrage orientiert sich am Konzept von *Müller et al.*, der folgende sechs Punkte beschreibt, die eine gute Umfrage ausmachen: 1. Definition der Forschungsziele (vgl. Kapitel 4), 2. Zielgruppenbestimmung, 3. Spezifizierung des Fragebogendesigns, 4. Überprüfung und Pretests, 5. Umsetzung und 6. Datenanalyse und Interpretation⁶⁸.

5.1 Zielgruppenbestimmung

In chemischen Laboratorien an einer Universität oder außeruniversitären Forschungseinrichtung ist das Spektrum der Synthese- und Messmethoden, Simulationen, etc., durch die Forschungsdaten erzeugt werden, sehr groß. Der Workflow

⁶³ Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V. et al. o.J.(2001).

⁶⁴ Zerr 2003.

⁶⁵ Zerr 2003.

⁶⁶ Zerr 2003.

⁶⁷ Müller et al. 2014.

⁶⁸ Müller et al. 2014.

in der Chemie verläuft häufig über mehrere Schritte, die in der Regel nicht von einer Person durchgeführt werden⁶⁹. Beispielsweise ist die Synthese eines gewünschten Produktes häufig mehrstufig. Das Produkt sowie die Zwischenprodukte werden mittels verschiedenster Analysemethoden charakterisiert. Dieser Workflow stellt häufig eine Zusammenarbeit von wissenschaftlichen und nicht-wissenschaftlichen Mitarbeitende mit unterschiedlicher Gewichtung dar. Diese Tatsache wurde bei der Zielgruppenbestimmung berücksichtigt. Darüber hinaus wurde bedacht, dass Bachelor- und Masterstudierende in der Chemie eine temporäre, praktische Arbeit leisten, d.h. sie sind in die Arbeitsgruppen und Instituten integriert und erzeugen mit Ihrer Arbeit ebenfalls Forschungsdaten.

Die primäre Zielgruppe der Umfrage wurde auf wissenschaftliche und nicht-wissenschaftliche Mitarbeitenden der chemischen Fakultäten an Universitäten und außeruniversitären Forschungseinrichtungen in Nordrhein-Westfalen (NRW) festgelegt.

Konkretisiert sind dieses: Professorinnen und Professoren, wissenschaftliche Mitarbeitende wie z.B. Doktorandinnen und Doktoranden, Postdocs, akademische Räte und technische Mitarbeitende wie z.B. Chemielaboranten und chemisch-technische Assistenten. Auch Bachelor- und Masterstudierende, die während der Umfrage Ihre Masterarbeit schreiben/ geschrieben haben, wurden mit einbezogen.

Eine Beschränkung der Zielgruppe auf NRW erfolgte aufgrund des zeitlichen Rahmens der Arbeit. In NRW sind an 12 Universitäten Fakultäten der Chemie angesiedelt, welche zirka ein Viertel aller Universitäten mit Fakultäten der Chemie deutschlandweit ausmachen⁷⁰. In NRW weisen folgende Universitäten Fakultäten der Chemie⁷¹ auf:

- Rheinisch-Westfälische Technische Hochschule Aachen
- Universität Bielefeld
- Ruhr-Universität Bochum
- Rheinische Friedrich-Wilhelms-Universität Bonn
- Technische Universität Dortmund
- Universität Duisburg-Essen

⁶⁹ Technische Informationsbibliothek et al. 2010.

⁷⁰ xStudy SE 1997-2019.

⁷¹ In alphabetischer Reihenfolge

- Heinrich-Heine-Universität Düsseldorf
- Universität zu Köln
- Westfälische Wilhelms-Universität Münster
- Universität Paderborn
- Universität Siegen
- Bergische Universität Wuppertal

Durch die Verknüpfung des *Forschungszentrums Jülich* mit Lehrstühlen an der *RWTH Aachen*, der *Universität Bonn*, der *HHU Düsseldorf* sowie der *Universität zu Köln* wird dieses mit in die Zielgruppe aufgenommen.

Mit der durchgeführten Umfrage in NRW wurden die differierenden Teildisziplinen wie Anorganische, Organische, Physikalische, Theoretische, Technische, Analytische, Makromolekulare, Lebensmittel-, Bio-, Bau- und Werkstoffchemie sowie die Didaktik der Chemie abgedeckt.

5.1.1 Einladung der Teilnehmenden und Rücklaufquote

Die Verteilung der Einladungen erfolgte auf zwei verschiedenen Wegen. Zunächst wurde die Ansprechperson der jeweiligen Fakultät kontaktiert und um eine Weiterleitung der Einladung gebeten. Dieses hat die Vorteile, dass alle Mitarbeitenden der jeweiligen Fakultät oder des jeweiligen Instituts zum gleichen Zeitpunkt durch eine interne Person erreicht wurden. Von Nachteil sind zum einen die mangelnde Bereitschaft der Ansprechpersonen sowie Leitlinien, die das Verteilen von Umfragen im Allgemeinen untersagen. Lediglich vier Ansprechpersonen haben die Umfrage über entsprechende Verteiler weitergeleitet. Die Fakultät für Chemie der *Universität Bielefeld* hat sich aus datenschutzrechtlichen Gründen gegen eine Teilnahme an der Umfrage ausgesprochen.

In einem weiteren Schritt wurden die Professorinnen und Professoren sowie die wissenschaftlichen und nicht-wissenschaftlichen Mitarbeitenden in einer aktiven Rekrutierungsform persönlich per Email angeschrieben. Anhand des zeitlichen Verlaufs nach Versenden der Emails ließ sich feststellen, dass das persönliche Anschreiben die Zielgruppe stärker zu einer Teilnahme motivierte als durch das Weiterleiten der Email durch eine vertraute Person. Das Vorgehen bei den Erinnerungsemails wurde entsprechend angepasst, so dass im Zeitrahmen von 4 Tagen bis 1 Tag vor Ende der

Umfrage auf das persönliche Anschreiben mit Ausnahme an der *RWTH Aachen* umgestellt wurde. Abbildung 3 ist zu entnehmen, dass durch das Versenden der Erinnerungsemail die Teilnahme deutlich gesteigert werden konnte.

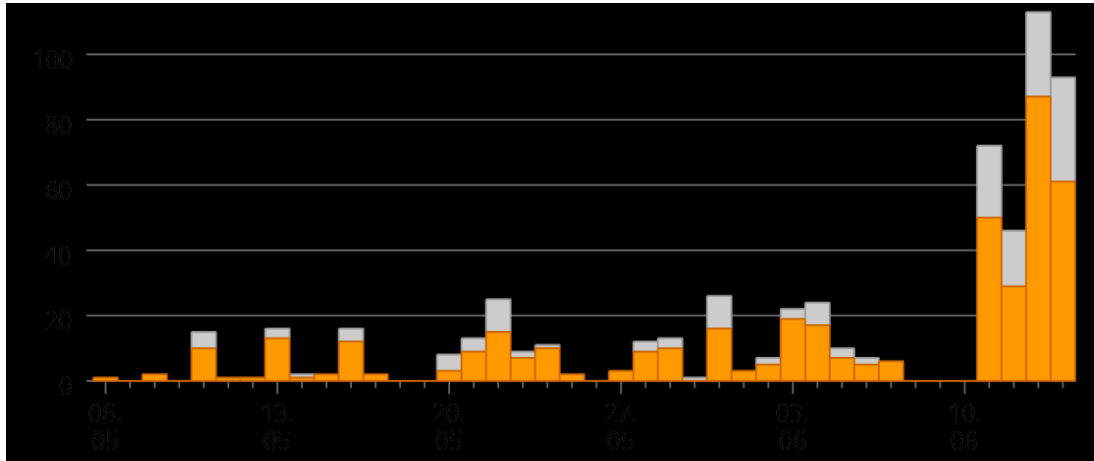


Abbildung 3: Rücklaufstatistik über den Zeitraum vom 06.05. bis 14.06.2019 (Quelle: www.soscisurvey.de)

Die Schwierigkeit, die sich hierbei stellte, war, dass nicht alle Emailadresse und teilweise auch keine Namen der Zielgruppenpersonen veröffentlicht sind. In solchen Fällen wurde die Professorin/ der Professor mit der Bitte zur Teilnahme und Weiterleitung der Umfrage an ihre/ seine Mitarbeitenden kontaktiert.

Wie bereits erwähnt, ist eine konkrete Bestimmung der Rücklaufquote aufgrund der fehlenden Grundgesamtheit nicht möglich, sondern kann nur über die Anzahl der versendeten Emails und der auf den Internetseiten der Fakultäten gelisteten Namen der Mitarbeitenden ohne E-Mailadresse angenähert werden.

Tabelle 1: Potentiell Teilnehmenden der Befragung

Anzahl versendete Einladungsemails	2966
Anzahl Beschäftigter laut Homepages (ohne E-Mailadressenangaben)	503
Unzustellbar	33
Abwesend	14
Summe	3422

In dem Erhebungszeitraum wurde der Fragebogen 610 Mal aufgerufen, wobei versehentliche Doppelklicks und Aufrufe durch Suchmaschinen inkludiert sind⁷². Insgesamt haben 584 Teilnehmende den Fragebogen ausgefüllt, wobei diesen 427 (73,12 %) abgeschlossen und 157 (26,88 %) abgebrochen haben (vgl. Abbildung 4). Als abgeschlossen werden alle Fragebögen gewertet, die Seite 13 als letzte Seite bearbeitet haben. Seite 14 enthält die Danksagung und Raum für Fragen und Kommentare (vgl. Abbildung 41).

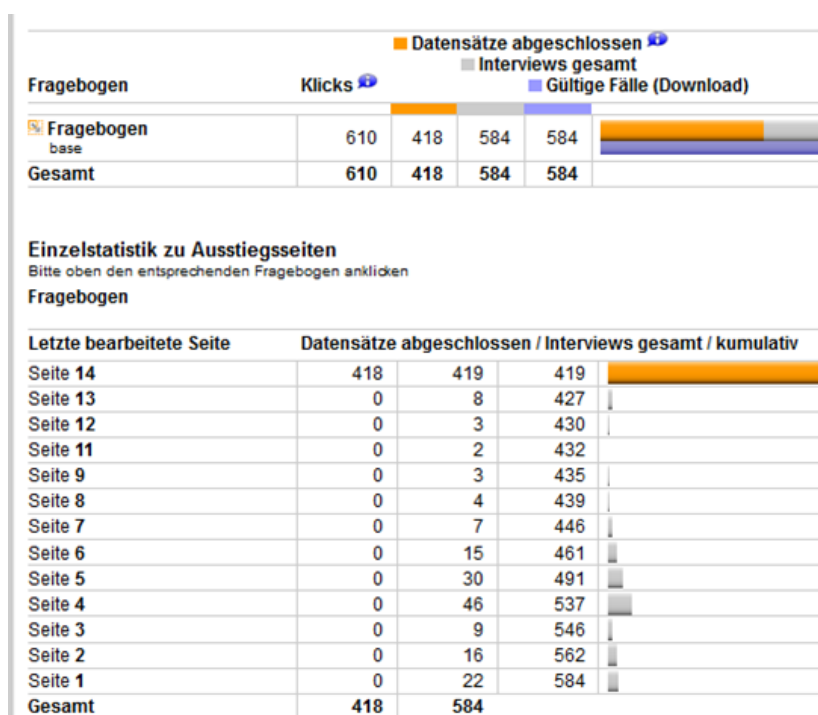


Abbildung 4: Screenshot der Datensätze und Abbrüche (Quelle: www.soscisurvey.de)

Aus den potentiell Teilnehmenden (vgl. Tabelle 1) und den vollständig oder partiell ausgefüllten Fragebögen lassen sich folgende Brutto-/Nettorücklaufquoten⁷³ schätzen:

$$\begin{aligned}
 \text{Bruttorücklaufquote} &= \frac{\text{Anzahl aller ausgefüllten Fragebögen}}{\text{Anzahl potentiell Teilnehmender}} * 100 \% \\
 &= \frac{584}{3422} * 100 \% = 17,06 \%
 \end{aligned}$$

⁷² SoSciSurvey GmbH o.J.

⁷³ www.wirtschaftslexikon24.com o.J..

Nettorücklaufquote

$$\begin{aligned} &= \frac{\text{Anzahl aller vollständig ausgefüllten Fragebögen}}{\text{Anzahl potentiell Teilnehmender}} * 100 \% \\ &= \frac{427}{3422} * 100 \% = 12,47 \% \end{aligned}$$

In der Literatur finden sich sehr unterschiedliche Angaben zu Rücklaufquoten bei Online-Umfragen. In vielen Beiträgen wie z.B. von *Wagner und Hering*⁷⁴ werden keine durchschnittlichen Rücklaufquoten aufgeführt, da Rücklaufquoten sehr stark von der Zielgruppe abhängen. Wenige Beispiele wie *Wilde* in einem Blogbeitrag⁷⁵ nennt zwischen 2,5 - 10 %, oder *Gräf* gibt 20 – 60 % bei Online-Kundenbefragungen an⁷⁶. Durch die sehr unterschiedlichen Angaben lässt sich keine qualitative Aussage über die erzielte Rücklaufquote treffen.

5.2 Spezifizierung des Fragebogens

Fragebögen, unabhängig vom jeweiligen Befragungszweck, sollten folgende Elemente enthalten⁷⁷:

5.2.1 Einleitung oder Einführungstext

Der Fragebogen wird als Online-Umfrage umgesetzt und die Zielgruppe wird per E-Mail auf die Umfrage aufmerksam gemacht. In der E-Mail wird ein Einführungstext zur Umfrage gegeben, welcher den Titel, den Grund, den Zeitraum und die durchschnittliche Bearbeitungszeit des Fragebogens sowie den Link und das Passwort zur Umfrage enthält.

5.2.2 Fragen und Antwortvorgaben sowie offene Texteingaben

Um eine möglichst hohe Vergleichbarkeit der Daten untereinander zu erzielen, wird der Fragebogen mit den Fragen, der Fragenreihenfolge sowie den Antwortvorgaben standardisiert. Offene Texteingabefelder sowie die Möglichkeit zur Mehrfachantwort werden bei den Fragen- und Antwortkonstruktionen berücksichtigt, so dass sich der Fragebogen aus offenen, geschlossenen und Hybridfragen aufbaut⁷⁸.

⁷⁴ Wagner und Hering 2014.

⁷⁵ Wilde o.J.

⁷⁶ Lorenz Gräf 2010.

⁷⁷ Fühles-Ubach 2013b.

⁷⁸ Schweiger o.J.

Der Fragebogen enthält sowohl Soziodemografische, wie auch Wissens- und Verhaltensfragen⁷⁹.

5.2.3 Konkrete und verständliche Hinweise

Bei der Fragen- und Antwortformulierung wurde darauf geachtet, dass konkrete und verständliche Hinweise eingearbeitet wurden und fachfremde Begriffe wie z.B. Metadaten erklärt wurden.

Abgeleitet aus den vorangestellten Überlegungen und den Pretests (vgl. Abschnitt 5.3) wurde ein Fragebogen mit 13 Fragen, der in acht Teilbereiche untergliedert wurde:

(a) Position der Teilnehmenden

Welche Position haben Sie an Ihrem Institut inne?

Diese Frage dient zum Einstieg und hilft bei der Ermittlung des Informationsbedarfs und –vermittlung der unterschiedlichen Gruppen der Teilnehmenden (Professorinnen und Professoren, wissenschaftliche Mitarbeitende und nicht-wissenschaftliche Mitarbeitende). Korrelationen zu nachfolgenden Fragen können aufgestellt werden.

(b) Ist-Zustand/ Wissen um die FAIR-Prinzipien

Kennen Sie die FAIR-Prinzipien? Ist Ihnen die folgende Grafik vertraut?

Das Ziel dieser Frage ist die Erhebung des aktuellen Wissensstands zu den FAIR-Prinzipien.

(c) Datenarten

1.) *Welche Daten erheben Sie bei Ihrer Forschung?*

2.) *Welche Versionen Ihrer Daten z.B. Rohdaten, aufgearbeitete, analysierte oder berechnete Daten speichern Sie am Projektende ab?*

Diese Fragen dienen dazu die Datenarten nach der Erhebungsmethode und nach den verschiedenen Versionen zu erfassen. Das Fragen nach dem Fachbereich war laut des Datenschützers der RWTH Aachen University nicht zulässig.

⁷⁹ Schweiger o.J.

(d) Auffindbarkeit

Um Daten innerhalb Ihres Instituts, Ihres Projektes, für eine Publikation, etc. auffindbar zu machen, ist die Beschreibung Ihrer Daten mit "weiterführenden Informationen" sogenannten Metadaten essentiell.

Wenn Sie nun an Ihre Daten denken. Mit welchen Metadaten beschreiben Sie Ihre Daten?
Bei diesem Block wird der Terminus Metadaten eingeführt. Die Bewertung einzelner Metadaten soll Aufschluss über die Bedeutung über die verschiedenen Teildisziplinen geben.

(e) Verständlichkeit der Daten und Laborbücher für den Datenzugriff

- 1.) Könnte eine Wissenschaftlerin/ ein Wissenschaftler innerhalb oder außerhalb Ihres Instituts mit den bisher von Ihnen mitgelieferten Metadaten, Ihre Daten verstehen?*
- 2.) Was für eine Art Laborbuch (Papier oder elektronisches Laborbuch) verwenden Sie? Welche Gründe sprechen für die jeweilige Nutzung?*

In diesem Block liegt der Fokus auf der Verständlichkeit aus Sicht der Teilnehmenden und auf dem am weitesten verbreiteten Instrument im chemischen Labor dem »Laborbuch«. Die Fragen nach dem physischen Zugang werden unter Punkt (7) aufgegriffen.

(f) Strukturierung der Daten und Vorgaben

- 1.) Neben den Metadaten ist auch die Strukturierung ihrer Daten ein wichtiges Element. Wie strukturieren bzw. benennen Sie ihre Daten?*
- 2.) Strukturieren und/ oder beschreiben Sie Ihre Daten mit Metadaten nach bestimmten Vorgaben, Absprachen, usw.?*

Dieser Block soll Aufschluss darüber geben, in wie weit Metadaten mit in die Strukturierung einfließen und ob Vorgaben zur Strukturierung gemacht werden.

(g) Datenweitergabe und Datenpublikation

- 1.) Wie teilen Sie Daten innerhalb Ihres Instituts?*
- 2.) Wie teilen Sie Daten außerhalb Ihres Instituts?*
- 3.) Haben Sie bereits Daten in Form von Rohdaten, bearbeiteten Daten, ... veröffentlicht?*
- 4.) Haben Sie selbst schonmal Daten aus anderen Quellen nachgenutzt?*

In diesem Block soll sowohl das interne und externe Datenteilen sowie das Datenveröffentlichen beleuchtet werden.

(h) Danksagungen und Anmerkungen

Zum Schluss erfolgt die Danksagung an der Teilnahme und den Teilnehmenden wird die Möglichkeit für Fragen und Kommentare gegeben.

5.3 Durchführung und Ergebnisse des Pretests

Im Vorfeld der Umfrage wurde ein Pretest durchgeführt, an dem vier Chemikerinnen und Chemiker sowie drei Vertreterinnen und Vertreter von Infrastruktureinrichtungen teilnahmen. Die Pretests wurden anhand des Textdokuments und des eingesetzten Tools SoSciSurvey durchgeführt. Durch die Pretests ließen sich Fragen und Antworten auf ihre Eindeutigkeit und Verständlichkeit überprüfen. Erklärungen bei einzelnen Fragen wurden ergänzt. Die Reihenfolge der Fragen wurde überarbeitet⁸⁰ und der Fragebogen wurde auf die vorangestellten 13 Fragen reduziert, in dem Fragen teilweise zu einer Frage zusammengefasst oder in Antworten verschiedener Fragen aufgegriffen wurden. Nicht zielführende oder redundante Fragen wurden entfernt⁸¹. Durch die Reduktion wurde eine durchschnittliche Bearbeitungsdauer von 12 Minuten erzielt, was im empfohlenen Zeitfenster von 10 bis 15 Minuten liegt⁸². Um die Verständlichkeit bei der Zielgruppe zu erhöhen, wurde ein Teil der Fragen in die »Sprache der Chemikerinnen und Chemiker« umformuliert⁸³.

6 Umsetzung

Die Online-Umfrage zur „FAIRness der Forschungsdaten in der Chemie“ wurde in dem Zeitraum vom 06. Mai 2019 bis zum 14. Juni 2019 mittels des Tools SoSciSurvey⁸⁴ durchgeführt.

Die Befragung wurde mit dem Datenschutzbeauftragten sowie mit dem wissenschaftlichen Personalrat der *RWTH Aachen University* abgestimmt.

⁸⁰ Mohler und Porst 1996.

⁸¹ Beywl und Schepp-Winter 2000.

⁸² Wagner und Hering 2014.

⁸³ Scheuch 1996.

⁸⁴ SoSciSurvey GmbH o.J.

Die Teilnehmenden wurden wie bereits erwähnt aktiv⁸⁵ durch einen Institutsverteiler oder durch eine personalisierte E-Mail rekrutiert. Die E-Mail enthielt den Link zur Umfrage sowie das Passwort, welches als Sicherheitsmaßnahme eingerichtet wurde.

Des Weiteren wurde auf das Layout geachtet, welches aus Vorlagen bei SoSciSurvey ausgesucht wurde. Die Fragen wurden so aufgeteilt, dass jeweils nur eine Frage bis maximal zwei Fragen gleichzeitig sichtbar waren. Dadurch ließ sich der Text der Umfrage auch auf einem Smartphone gut lesen. Um für alle Beteiligten möglichst die gleichen Voraussetzungen zu schaffen, wurde die Umfrage auf den gängigsten Browsern *Mozilla Firefox*, *Microsoft Edge* und *Google Chrome* getestet⁸⁶.

Ein »Zurück-Button« wurde eingefügt, um die Möglichkeit zur Korrektur bereits gegebene Antworten zu offerieren⁸⁷.

6.1 SoSciSurvey

Eine Umsetzung der Umfrage erfolgte mit dem Tool *SoSciSurvey*, da dieses kostenlos für nicht-kommerzielle Zwecke zur Verfügung steht. Es ist barrierefrei und datenschutzkonform laut DSGVO. Ein weiter ausschlaggebender Punkt war, dass der Befragungsserver in München steht, wo auch die Betreiber ihren Sitz haben. Somit ist sichergestellt, dass die Daten dem strengen europäischen Datenschutzgesetz unterliegen. Außerdem überzeugte *SoSciSurvey* mit dem Angebot von unterschiedlichen Fragetypen und der Einbindung von Bilddateien in einzelnen Fragen. Die Umfrage kann jederzeit online verfolgt werden und die Daten können in den verschiedensten Formaten ausgegeben werden⁸⁸.

7. Auswertung der erhobenen Daten

Vor der Datenauswertung und Interpretation erfolgt der Schritt der Datenaufbereitung.

Von den 584 erfassten Datensätzen werden lediglich die vollständig ausgefüllten Fragebögen berücksichtigt. Dieses bedeutet, dass alle 13 Fragen bearbeitet wurden, auch wenn manche übersprungen wurden. Fragebögen, bei denen einzelne Fragen

⁸⁵ Thielsch und Weltzin 2012.

⁸⁶ Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V. et al. o.J.(2001).

⁸⁷ Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V. et al. o.J.

⁸⁸ SoSciSurvey GmbH o.J.

übersprungen werden, deuten auf ein gewisses Interesse der Teilnehmenden. Es kann davon ausgegangen werden, dass die Fragen ehrlich beantwortet wurden und werden aus diesem Grund in die Stichprobe mit einbezogen. Aus beiden Kriterien ergeben sich zusammen 428 Datensätze. Anschließend werden Teilnehmende mit Scherzantworten herausgefiltert, wodurch 427 Datensätze verbleiben. Eine Aufbereitung nach der Verweildauer erfolgte nicht, da die Reaktionszeit laut dem Datenschutzbeauftragten der RWTH Aachen University, Herrn *Löbner* nicht erfasst werden durfte⁸⁹. Daher wird für die Auswertung eine Stichprobe $n = 427$ herangezogen (vgl. Kapitel 5.1.1).

Da es sich um eine quantitative Erhebung handelt, werden die Daten mittels deskriptiver Statistik ausgewertet. Die erhobenen Daten werden zusammengefasst und mittels Tabellen und/ oder Diagrammen visualisiert. Absolute und relative Häufigkeiten⁹⁰ bestimmter Merkmalausprägungen und Kennzahlen wie z.B. der Mittelwert und Merian untermauern das Ergebnis.⁹¹ Bei Mehrfachantworten bezieht sich die relative Häufigkeit auf die Anzahl der Antworten.

Die Datenauswertung erfolgt mit Microsoft Office Excel 2016.

Die ausgewerteten Daten sowie die Rohdaten sind auf Zenodo: DOI 10.5281/zenodo.3415059 unter einer CC-BY_NC 4.0 Lizenz veröffentlicht.

7.1 Position der Teilnehmenden

Als Einstiegsfrage *Welche Position haben Sie an Ihrem Institut inne?* wurde eine soziodemografische Frage gewählt. Bei dieser Frage handelt es sich um eine Einfachauswahl und es ist eine offene Texteingabe unter »Andere« möglich. Die Teilnehmenden mussten diese Frage beantworten. An der Umfrage haben insgesamt 47 Professorinnen und Professoren (Privatdozentinnen und Privatdozenten inkludiert) (11 %) teilgenommen. Aus der Gruppe der wissenschaftlichen Mitarbeitenden haben 243 Doktorandinnen und Doktoranden (57%) sowie 76 Senior Scientists bestehend aus Postdocs, Akademische Rätin, Akademischer Rat, Oberingenieurin, Oberingenieur, Habilitandin, Habilitand, Nachwuchsgruppenleiterin und Nachwuchsgruppenleiter

⁸⁹ Helmut Löbner 2019.

⁹⁰ In Klammern nach dem jeweiligen Merkmal eingefügt nach dem Schema: (absolute Häufigkeit; relative Häufigkeit in Prozent).

⁹¹ Wendler 2018.

(18 %) teilgenommen. Des Weiteren haben 37 technische Mitarbeiterinnen und Mitarbeiter (9 %) sowie 12 Master- (3 %) und 2 Bachelorstudierende (0,5 %) die Umfrage bis zum Ende abgeschlossen. Zehn weitere Teilnehmende (2 %) konnten keiner der voran genannten Gruppen zugeordnet werden, da die Angaben wie Student, studentische oder wissenschaftliche Hilfskraft oder abgeordnete Lehrkraft zu allgemein waren. (vgl. Abbildung 5 **Fehler! Verweisquelle konnte nicht gefunden werden.**).

Position der Teilnehmenden

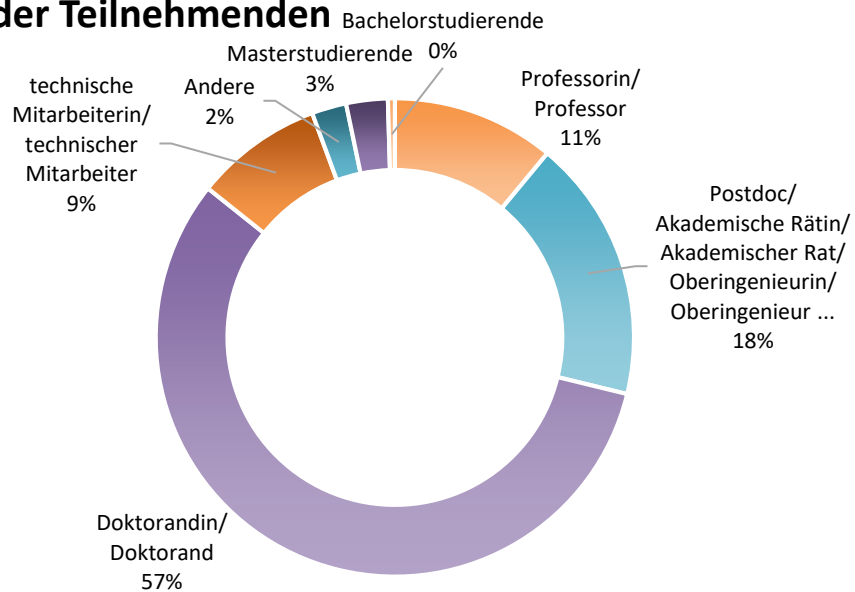


Abbildung 5: Auswertung der Frage - Welche Position haben Sie an Ihrem Institut inne?

Wie bereits erwähnt, ließ sich kein Vergleich zur Grundgesamtheit aufstellen, da teilweise eine Aufschlüsselung der Position der jeweiligen Person auf den Homepages fehlt. Stichprobenmäßig ließ sich aus den veröffentlichten Daten als Trend das Verhältnis Professorin/ Professor: wissenschaftliche Mitarbeitende: nicht-wissenschaftliche Mitarbeitende = 1:8:1 ableiten. Die Umfrage ausgefüllt haben 1:7:1. Es zeigt sich, dass verhältnismäßig weniger wissenschaftliche Mitarbeitende an der Umfrage

teilgenommen haben als Professorinnen und Professoren.

Zusammenfassend lässt sich sagen, dass die Beteiligung an der Online-Umfrage in einem Zeitraum über 6 Wochen bei 12 % lag, was bestätigt, dass es ein gewisses Interesse am Thema Forschungsdatenmanagement gibt. Verhältnismäßig haben mehr Professorinnen und Professoren an der Umfrage teilgenommen als wissenschaftliche Mitarbeitende.

7.2 Ist-Zustand

Mit der Frage *Kennen Sie die FAIR-Prinzipien? Ist Ihnen die folgende Grafik vertraut?* soll der aktuelle Stand des Wissens um die FAIR-Prinzipien in der Chemie ermittelt werden. In der Frage ist die FAIR-Grafik eingebunden und über eine Dropdown-Auswahl stehen vier verschiedene Antwortoptionen als Einfachauswahl zur Verfügung. Drei Antwortoptionen sind in gestaffelter Weise positiv und eine ist negativ. Die Frage musste durch die Teilnehmenden beantwortet werden.

Wissen um die FAIR-Prinzipien

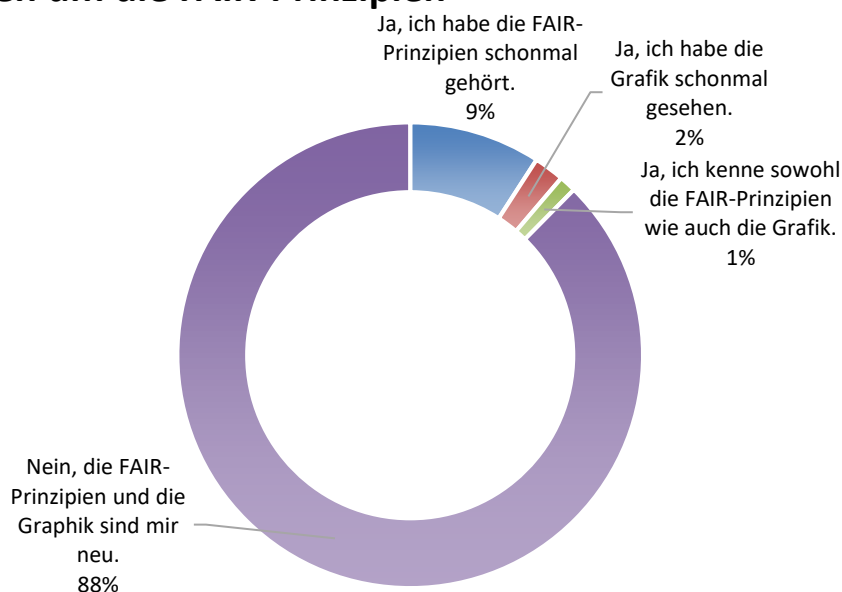


Abbildung 6: Auswertung der Frage - *Kennen Sie die FAIR-Prinzipien? Ist Ihnen die folgende Grafik vertraut?*

Die Ergebnisse zeigen, dass so sehr Infrastruktureinrichtungen mit den FAIR-Prinzipien vertraut sind und die FAIR-Grafik (vgl. Abbildung 2) für die unterschiedlichsten Zwecke eingesetzt wird, kennt nur knapp jeder 8. Teilnehmende aus der Chemie die FAIR-Prinzipien. Nur 39 Teilnehmende (9 %) haben von den FAIR-Prinzipien gehört.

9 Teilnehmende (2 %) haben die FAIR-Grafik gesehen und 5 weitere Teilnehmende (1 %) stimmten beiden zu. Einer deutlichen Mehrheit von 374 Teilnehmenden (88 %) sind die FAIR-Prinzipien unbekannt. *Howes* macht mit ihrem Beitrag „*Chemistry data should be FAIR, proponents say. But getting there will be a long road*“ auf diese Problematik aufmerksam. Das Arbeiten und die Datenerzeugung haben sich in der Chemie in den letzten 50 Jahren sehr verändert, nur haben sich einzelne Umgebungsbedingungen nicht mitentwickelt.⁹²

Werden die 53 Teilnehmenden, die die FAIR-Prinzipien/ -Grafik kennen, hinsichtlich ihrer Position betrachtet, zeigt die Korrelation eine Abnahme des Wissens mit der Position (vgl. Tabelle 2).

Tabelle 2: Korrelation Position zum Wissen um die FAIR-Prinzipien/ -Grafik

Position	Summe der teilnehmenden Gruppe	Wissen um die FAIR-Prinzipien/ -Grafik	Verhältnis
Professorin/ Professor	47	12	ca. 4:1
Wissenschaftliche Mitarbeitende	319	33	ca. 10:1
Nicht-wissenschaftliche Mitarbeitende	37	3	ca. 12:1
Masterstudierende	12	4	3:1
Andere	10	1	10:1

Jede 4. Professorin/ jeder 4. Professor kennt die FAIR-Prinzipien/ -Grafik, während es bei den wissenschaftlichen Mitarbeitenden sowie bei Andere nur noch jeder 10. und bei den nicht-wissenschaftlichen Mitarbeitenden nur noch jeder 12. ist. Am besten ist das Verhältnis mit 3:1 bei den Masterstudierenden. Eine mögliche Begründung könnte sein, dass sie bei den ersten Recherchen zur Masterarbeit auf Webseiten von Infrastruktureinrichtungen, der RDA oder Go FAIR stoßen und auf das Thema aufmerksam werden. Professorinnen und Professoren besuchen im Gegensatz zu

⁹² Howes 2019.

wissenschaftlichen und nicht-wissenschaftlichen Mitarbeitenden auch Veranstaltungen mit politischem Charakter, wo mit hoher Wahrscheinlichkeit FDM und FAIR thematisiert werden. Auf dem Wissenschaftsforum 2019 in Aachen, dem wichtigsten Chemiekongress im deutschsprachigen Raum, findet zum ersten Mal eine Veranstaltung „Forschungsdaten und digitaler Wandel“ organisiert durch die Initiative NFDI4Chem statt. Zusätzlich wird es einen Stand geben, bei dem alle Teilnehmende sich über die Initiative und alle FDM relevanten Themen informieren können.

Zusammenfassend lässt sich sagen, dass die FAIR-Prinzipien dem größten Teil der Teilnehmenden (88 %) unbekannt sind. Betrachtet man die Korrelation Wissen FAIR-Prinzipien zu den Positionen der Teilnehmenden, zeigt sich, dass Masterstudierende und Professorinnen und Professoren am besten über die FAIR-Prinzipien informiert sind. Die Initiative NFDI4Chem informiert innerhalb der Chemie zu FDM relevanten Themen.

7.3 Datenarten

Die erste Frage in diesem Block nach den Datenarten *Welche Daten erheben Sie bei Ihrer Forschung?* soll die teildisziplinspezifische chemische Abdeckung widerspiegeln. Die direkte Frage nach der Zugehörigkeit zu einzelnen Teildisziplin wurde als datenschutzrechtlich schwierig eingestuft, da einzelne Teildisziplinen sehr klein sind und Antworten zugeordnet werden könnten. Daher wurde die Frage nach den Datenarten entwickelt. Bei den Datenarten mit dem Merkmal »Methode« weisen einzelne Teildisziplinen untereinander Überschneidungen auf. Durch die Verwendung der gleichen Methode haben die Teilnehmenden eine gemeinsame Basis unabhängig von der Teildisziplin.

Die Frage ist eine Hybridfrage mit vorgegebenen Antworten zur Mehrfachauswahl sowie der Möglichkeit zur offenen Texteingabe. Ein Überspringen der Frage war möglich. Insgesamt wurden 1521 Antworten gegeben. Es zeigt sich, dass die

Anzahl gewählter Antwortoptionen

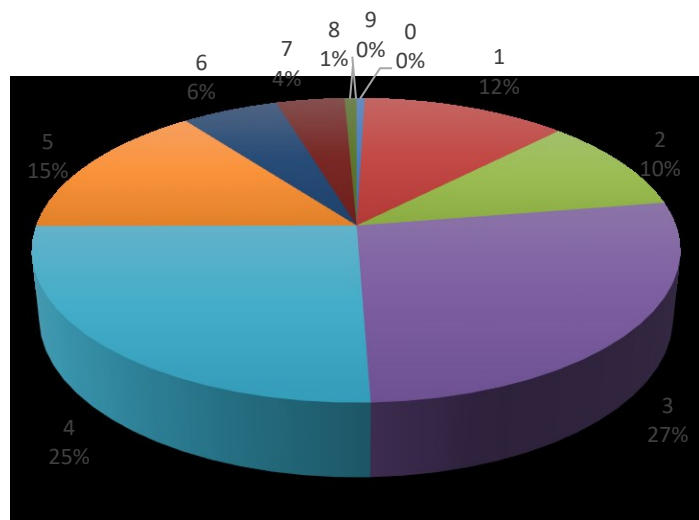


Abbildung 7: Auswertung der Frage - „Welche Daten erheben Sie bei Ihrer Forschung?“; Anzahl gewählter Antwortoptionen & relative Häufigkeit der Teilnehmenden

Teilnehmenden im Schnitt mit 3,6 verschiedenen Datenarten arbeiten (vgl. Abbildung 7). Innerhalb der Datenarten entstehen die unterschiedlichen Dateiformate wie z.B. es bei der NMR-Spektroskopie in der Regel ein proprietäres Format des Geräteherstellers ist und bei der UV-Vis-Spektroskopie die Daten als csv-File ausgegeben werden. Es lässt sich folgern, dass die Teilnehmenden durchschnittlich in ihrem Alltag mit $\geq 3,6$ Dateiformaten umgehen müssen.

Datenarten mit dem Merkmal »Methode«

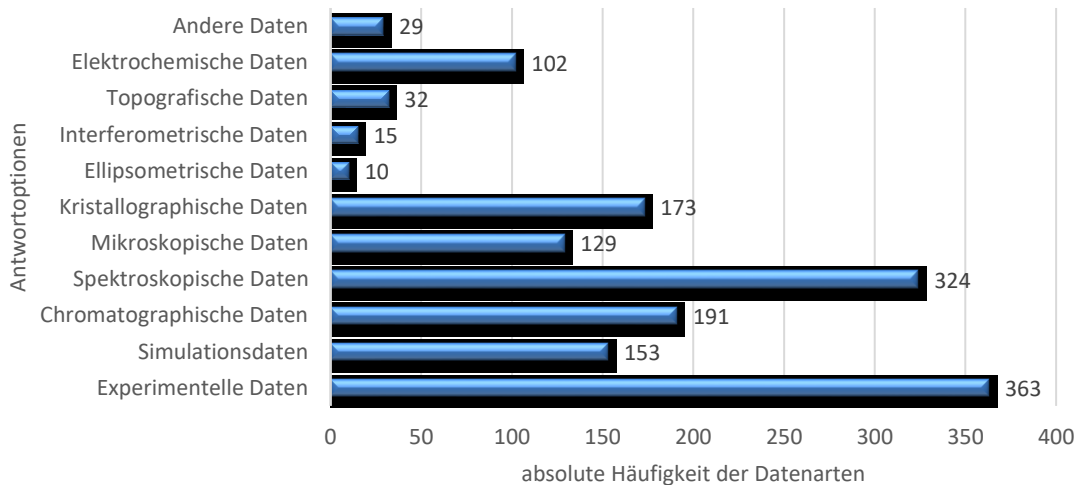


Abbildung 8: Auswertung der Frage – Welche Daten erheben sie bei Ihrer Forschung?

Am häufigsten, 363-mal (24 %) sind die experimentellen Daten genannt worden (vgl. Abbildung 8). Laut *Duden* versteht man unter einem Experiment „*einen Versuch, durch den etwas entdeckt, bestätigt oder gezeigt werden soll*“.⁹³ Nach dieser Definition werden in allen Teildisziplinen experimentelle Daten erzeugt. Doch wo beginnt ein Experiment und wo endet es? Beispielweise sei ein chemisches Experiment in der Polymerchemie betrachtet. Klassischerweise wird die meist mehrstufige Synthese des gewünschten Produktes als Experiment bezeichnet, bei dem unterschiedliche experimentelle Daten wie z.B. Farbe, Löslichkeit und Aggregatzustand anfallen. Die anschließenden analytischen Methoden wie die NMR-Spektroskopie, die Gelpermeationschromatographie (GPC) oder die Elementaranalyse zur Strukturaufklärung und Charakterisierung der synthetisierten Substanz werden oft mit dem Terminus »Messung« beschrieben und nicht als experimentelle Daten wahrgenommen. Als weiteres Beispiel sei die Rasterkraftmikroskopie (AFM) aufgeführt, bei der atomare Kräfte im Nanometerbereich gemessen werden. Das Experiment beginnt mit der Präparation der Probe, gefolgt vom Einbau der Probe in das Gerät, dem Einstellen der Messparameter und dem Vermessen der Probe. Auch bei diesem Beispiel fallen experimentelle Daten an, nur werden diese häufig nicht als solche beschrieben. Ähnliches trifft auf die Simulationsexperimente zu, die ebenfalls zu den experimentellen Daten gehören, aber häufig als »Berechnungen« bezeichnet werden. In der Umfrage wurden sie 153-mal (10 %) gewählt. Mit diesen Beispielen soll gezeigt werden, dass der Bezug für experimentelle Daten unterschiedlich gesetzt wird und deshalb die relative Häufigkeit der gegebenen Antwort nicht bei 100 % liegt.

Die Datenarten, die bei analytischen Methoden entstehen, werden in den Antwortoptionen in die am weitesten verbreiteten Methodenarten aufgegliedert. Die am häufigsten verwendete Methode ist mit 324-facher Nennung (21 %) die Spektroskopie. Es folgen die Chromatographie (191; 13 %), die Kristallographie (173; 11 %), die Mikroskopie (129; 8 %) und die Elektrochemie (102; 7 %). Deutlich weniger wurden topographische Daten (32; 2%), interferometrische (15; 1 %) und ellipsometrische (10; 1 %) Daten genannt (vgl. Abbildung 8: Auswertung der Frage – Welche Daten erheben sie bei Ihrer Forschung? Abbildung 8). Unter »andere Daten« wurden 5-mal

⁹³ Duden o.J.

die massenspektrometrischen und 2-mal allgemein spektrometrische Daten genannt. Alle weiteren Datenarten wie molekularbiologische, nuklearmedizinische, biochemische, audio-visuelle oder chemiedidaktische Daten, um einige aufzuzählen, wurden einfach genannt.

Die zweite Frage *Welche Versionen Ihrer Daten z.B. Rohdaten, aufgearbeitete, analysierte oder berechnete Daten speichern Sie am Projektende ab?* in diesem Block zielt auf Datenarten mit dem Merkmal »Versionierung« ab. Die Frage ist als Hybridfrage mit Mehrfachauswahl und der Möglichkeit zur offenen Texteingabe konzipiert. Die

Anzahl gewählter Antwortoptionen

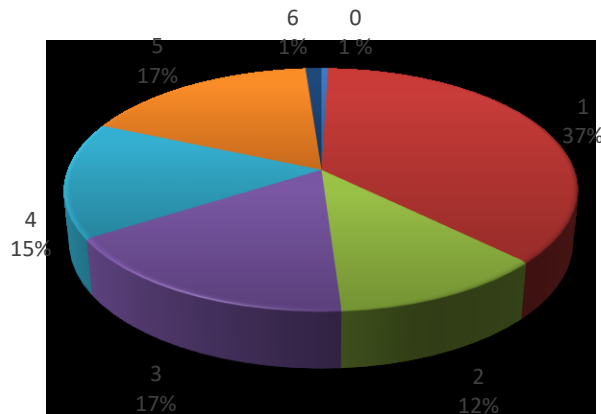


Abbildung 9: Auswertung der Frage - „Welche Versionen ihrer Daten speichern Sie am Projektende ab?“; Anzahl gewählter Antwortoptionen & relative Häufigkeit der Teilnehmenden

Frage konnte übersprungen werden. Insgesamt wurden 1140 Antworten optiert, wobei durchschnittlich 2,7 Antworten gegeben wurden (vgl. Abbildung 9).

Bei der Frage wird für die Versionierung unterschieden zwischen den Roh- (237; 21 %), aufgearbeiteten (218; 19 %), analysierten (225; 20 %) und berechneten Daten (162; 14 %). Als weitere Antwortoptionen wurden alle Daten (255; 22 %), das bedachte Löschen von Daten (43; 4 %) und andere Daten (0; 0 %) angeboten (vgl. Abbildung 10).

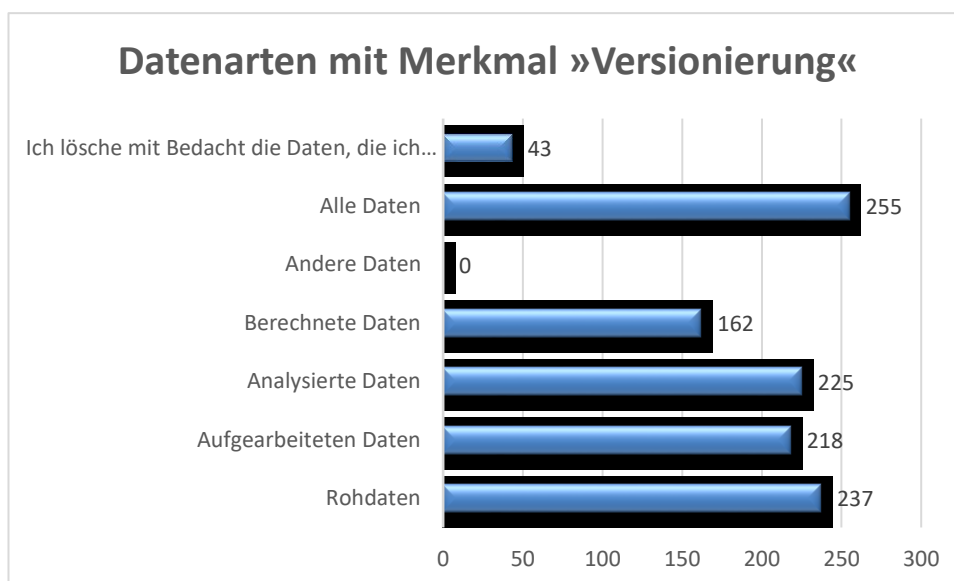


Abbildung 10: Auswertung der Frage DA02- verschiedene Datenarten zur Aufbewahrung

Das Maximum mit 255 Antworten (22 %) bildet das Speichern aller Daten. Hierdurch sowie durch das Minimum beim bedachten Löschen von Daten⁹⁴ wird gut die Unsicherheit bei der Auswahl der Daten zur Aufbewahrung deutlich widerspiegelt. Das Vorhandensein dieser Unsicherheit bei der Datenauswahl zur Aufbewahrung wurde in eigenständig von der Autorin durchgeführten Seminaren und Beratungsgesprächen zum FDM bei Forschenden beobachtet. „*Bevor ich ausversehen etwas Falsches lösche, bewahre ich lieber alles auf.*“ (Zitat eines Teilnehmenden des FDM I-Seminars an der RWTH Aachen University im Herbst 2018). Generelle Checklisten wie z.B. »5-steps to decide what data to keep«⁹⁵ oder Datenmanagementpläne (DMPs) unterstützen Forschende mit Kriterien bei der Auswahl ihrer Daten. Diese Checklisten beinhalten allgemeine Kriterien, die chemiespezifisch oder sogar für einzelne Teildisziplinen

⁹⁴ Die Antwortoption „andere Daten“ außen vorgelassen.

⁹⁵ Digital Curation Center 2014.

angepasst, erweitert oder gekürzt werden könnten, um so die Teilnehmenden besser zu unterstützen und die Qualität der aufbewahrten Daten im Durchschnitt zu steigern.

Zusammenfassend lässt sich sagen, dass in den chemischen Instituten an nordrhein-westfälischen Hochschulen am meisten experimentelle Daten (24 %) sowie spektroskopische Daten (21 %) anfallen. Diese werden in ihren unterschiedlichen Versionen alle gespeichert (22 %). Ein bewusstes Löschen der Daten (4 %) findet in geringem Maße statt. Die Entwicklung und anschließende Anwendung von chemiespezifischen Kriterien zur Aufbewahrung von Forschungsdaten würde durchschnittlich die Qualität der gespeicherten Daten erhöhen.

7.4 Auffindbarkeit

Die Auffindbarkeit der Daten ist ein wichtiger Teil der FAIR-Prinzipien, in denen den Metadaten eine besondere Bedeutung zukommt. Analysiert man die FAIR-Prinzipien, fällt auf, dass in 13 der 15 genannten Prinzipien die Metadaten fokussiert werden. Allerdings ist der Terminus Metadaten vielen Forschenden nicht geläufig⁹⁶, daher wird dieser in der Fragestellung als »weiterführende Informationen« eingeführt. Die Erfassung von und der Umgang mit Daten (und Metadaten) stellen den Alltag der Forschenden dar und sollten ihnen daher am vertrautesten sein. Die umfangreichen Metadaten wurden in diesem Frageblock in den Fokus gerückt.

Es wurde eine Item-Batterie-Frage entwickelt, die drei Metadaten (Strukturformel, Summenformel und Analyseverfahren) vorgibt und zehn Felder zur freien Texteingabe vorsieht. Des Weiteren konnten alle Metadaten auf einer Skala von 0 bis 9 (0 = unbedeutend bis 9 = sehr wichtig) bewertet werden. Die Frage *Mit welchen Metadaten beschreiben Sie Ihre Daten?* musste nicht beantwortet werden. Wird die

⁹⁶ Hausen und Windeck 2018..

Anzahl der angegebenen Metadaten betrachtet (vgl. Abbildung 11), lässt sich erkennen, dass 184 Teilnehmende (43 %) ein Metadatum aufgelistet haben. Neun und zehn Metadaten haben zwei Teilnehmende genannt. Es lässt sich schlussfolgern, dass den Teilnehmenden im Durchschnitt 2,1

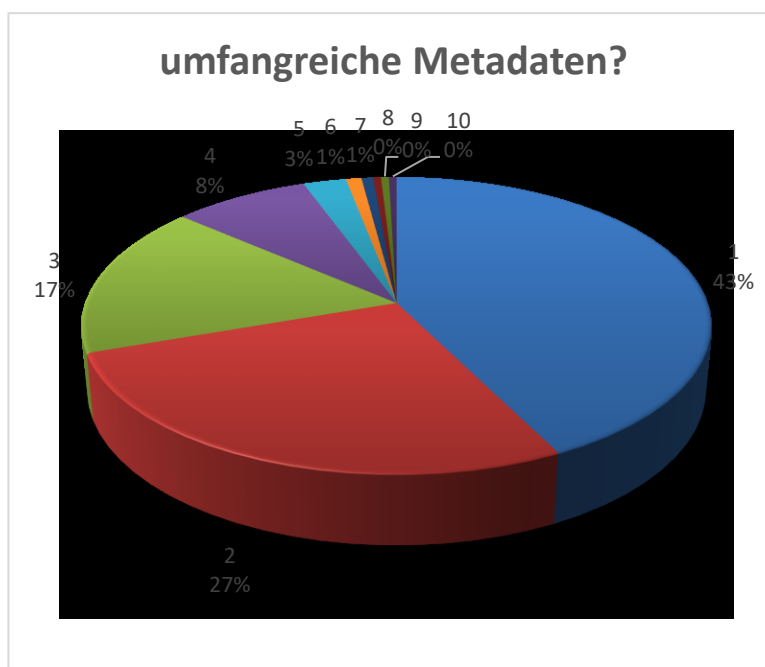


Abbildung 11: Auswertung der Frage - „Mit welchen Metadaten beschreiben Sie Ihre Daten?“; Anzahl gegebener Metadaten & relative Häufigkeit der Teilnehmenden

Metadaten zum Auffinden Ihrer Daten genügen. Eine Möglichkeit für die zurückhaltende Nennung der Metadaten kann die Unsicherheit sein, zu viel von den eigenen Daten Preis zu geben⁹⁷.

Insgesamt haben die Teilnehmenden 397 Metadaten eingetragen und bewertet. Es wurden 227 verschiedene Begriffe genannt, die teilweise mittels terminologischer Kontrolle zusammengefasst wurden. Als Beispiele sind hier zu nennen:

- 1.) experimentelle Parameter = Messparameter, Messkonditionen, verwendete Parameter, Bedingungen beim Experiment, Experimentelle Bedingungen, experimentelle Details, Parameter während der Messung
- 2.) Probenbezeichnung = Probenname, Probenbenennung, Probenkennung, Substanzname, Verbindungsname, Probenkürzel, Substanzkürzel, Versuchskürzel, eindeutige Probenbezeichnung (des Auftraggebers)
- 3.) Bearbeiter = Experimentator, Person, Name, Erzeuger der Daten, Name der bearbeitenden Person, Synthetiker der Probe, Ausführer.

⁹⁷ Bericht aus von der Autorin eigen gemachten Erfahrung während Seminaren und Beratungen.

Insgesamt entstand eine Liste von 176 verschiedenen Metadaten⁹⁸, die sich in drei Arten^{99, 100} und vier Unterarten kategorisieren ließen:

a) Administrative Metadaten:

Bearbeiter (10), Projekt (7), Studierende (1), Nutzer (1), Ablageort Rohdaten (1), Ordner auf dem Computer (1), Datum (36), Datum Datenerzeugung (1), Datum des Experiments (2), Datum Herstellung (1), Datum der Messung (4), Tag der Probenpräparation (1), Uhrzeit (6), Zeitangaben (3), Messort (2), (interne) Abkürzungen (3), Initialien (4), Interne Nummerierung der untersuchten Struktur (1), Laborjournal Kürzel (7), mp-ID (1), Nummer (2), Probe (8), Probenbezeichnung (17), Probennummer (18), Strukturname (2), Trivialname (1), IUPAC-Name (1), CAS-Nr. (1)

b) Relationship-Metadaten:

Zuordnung der Daten (1), Zuordnungslisten (1), Verknüpfung zur Veröffentlichung (1), Vergleichsdaten (1), Ursprung (1), Status der Weiterverarbeitung (1)

c) Inhaltlich-beschreibende bzw. chemiespezifische Metadaten:

Die fachlichen Metadaten lassen sich weiteraufgliedern. Wird ein Experiment, eine Messung oder eine Simulation wie ein Prozess betrachtet, erfolgt eine Zuordnung der Metadaten nach dem Eingabe-Verarbeitung-Ausgabe-Prinzip (EVA)¹⁰¹ (vgl. Abbildung 12). Auffällig ist die differierende Granularität der Metadaten, so sind 116 Metadaten wie Anregungswellenlänge, Lösungsmittel, ... konkret benannt und 12 Metadaten wie Ergebnis, Auswertung, ... allgemein formuliert. Bei 10 Metadaten ist keine eindeutige Zuordnung nach EVA in diesem Zusammenhang geben, da Metadaten wie Partikelgröße, ... für den einen Prozess als Eingabe und für den anderen als Ausgabe zählen. In Kombination mit weiteren Metadaten wie dem jeweiligen Experiment könnte dieses genauer aufgeschlüsselt werden. Diese Metadaten werden in diesem Fall als Eingabe betrachtet.

⁹⁸ In Klammern ist die absolute Häufigkeit der Metadatenennung angegeben.

⁹⁹ forschungsdaten.info o.J.

¹⁰⁰ Rühle 2012.

¹⁰¹ Wikipedia 2019a.



Abbildung 12: EVA-Prinzip

i.) Metadaten der Eingabe:

Analyseprogramm (1), Anregungsenergie (2), Anregungswellenlänge (1), Basissatz (4), Belichtungszeit (2), Beschichtungsmaterial (1), Detektionswellenlänge (1), Einwaagen (2), Elektrolyt (1), Enzym (1), experimentelle Parameter (16), Expressionsplasmid (1), Expressionsstamm (1), Farbstoff (1), Filter (1), Fitfunktion (1), Gaszusammensetzung (1), Gegenion (1), Geräteeinstellungen (1), Gitterparameter (2), Integrationszeit (1), Kalibration (1), Komponenten (1), Konfigurationsdateien (1), Laserintensität (1), Laserleistung (1), Lösungsmittel (5), Material (2), Messdauer (1), Messfühler (1), Messgerät (3), (Mess-)Temperatur (11), Messmodi (1), Mischungsverhältnis (1), Modelldetails (2), Molmasse (1), Name der Zelllinie (1), Ofenprogramm (1), Parameter (4), Partikelformen (1), Partikelgröße (1), pH-Wert (1), Pixelgröße (1), Probenkonzentration (1), Probenvorbereitung (5), Programmcode (1), Programmversionen (1), Proteinname (1), Proteinvariante (1), Raumgruppe (3), Reaktionsbedingungen (3), Rotationskonstanten (1), Setup (1), Simulationsparameter (7), Sintertemperatur (1), Sinterzeit (1), Software (1), Spektralbereich (1), Startgeometrien (1), Stoffgruppe (2), Substrat (1), Supramolekulares Model (1), Systemzusammensetzung (1), Umgebungsbedingungen (1), Wellenfunktionsansatz (1), Wellenlänge (1), Zelltyp (1), Zusammensetzung (3)

ii.) Metadaten der Verarbeitung:

Circulardichroismus (1), Cyclovoltammetrie (2), DFT-Methode (1), Elementaranalyse (1), Gelpermeationschromatographie (1) High Performance Liquid Chromatographie (1), Infrarotspektroskopie (3), Massenspektrometrie (7), Kernresonanzspektroskopie (9), Optimierungsmethode (1), TEM (1), TGA (1), UV-Vis-Spektroskopie (1), XPS (1), X-Ray (1), XRD (1), Berechnungsmethode (8), LVE (1), Schaumtest (1), Korrosionstest (1), Methode (6), Physisorption (1), Experiment (1), Experimenteller Aufbau (2)

iii.) Metadaten der Ausgabe:

Animationen (1), Ausbeuten (1), Auswerteergebnis (2), berechnete Energien (1), berechnete Spektren (1), Berechnungen (2), Bindungsmotiv (1), CAD-Zeichnungen (1), Chromatogramme (1), Diagramme (2), Ergebnisse (1), Ersatzschaltbild (1), Fluoreszenz (1), Fotos (1), gemessene Spektren (1), Kartesischen Koordinaten (1), Kristallstruktur (4), Kinetiken (3), Molekülgeometrie (1), Name Analyt (1), Numerische Genauigkeit (1), Optische Dichte (1), Output Quantenchemischer Rechnungen (1), Produkte der Lernenden (1), Ramanspektren (1), Rasterkraftmikroskopie-Bild (1), Reinheit (1), R_f -Wert (1), Schubspannung (1), Spektren (2), spektroskopische Fingerprints (1), Spinmultiplizität (1), Statistische Auswertung (1), Stromdichte (1), Ströme (1), Tabellenkalkulationen (2), TOF/ TON-Werte (1)

iv.) allgemein-chemiespezifische Metadaten:

Funktional (1), Hersteller (2), Korrekturen (1), Lagerung der Proben (1), Literaturzitate (3), Reaktionsgleichung (1), Reaktionstyp (1), Referenzen (1), Schlagwort (2), Skizzen (2), Zusammenfassung der Messergebnisse (1).

Von den 176 genannten Metadaten wurden 147 Metadaten ≤ 3 und 29 Metadaten ≥ 3 aufgezählt. Dieses ist ein Beispiel für die Vielfältigkeit und die Vielzahl der Metadaten in der Chemie. Die Prozessbetrachtung hilft einzelnen Forschenden, Arbeitsgruppen oder Instituten bei der Ermittlung von Metadaten. »Wo fallen Metadaten an? Wo werden Metadaten eingegeben? Und wo aus? Gehen Metadaten während des Prozesses verloren?«. Sie kann ebenfalls bei der Erarbeitung von Minimum of Information (vgl. Kapitel 2) und Metadatenschemata für eine konkrete Methode, einen konkreten Workflow oder einer Experimentenreihe unterstützen.

Die Metadaten sollten zusätzlich zur Nennung bewertet werden. Bei der Auswertung dieser Frage wurden die vorgegebenen Begriffe: Strukturformel, Summenformel und Analyseverfahren sowie die Metadaten mit ≥ 3 -Nennung berücksichtigt (vgl. Tabelle 3).

Forschungsdaten in der Chemie

Tabelle 3: Bewertung der genannten Metadaten von 0 bis 9 (0 = unbedeutend und 9 = sehr wichtig)

Bewertung	-9	-1	0	1	2	3	4	5	6	7	8	9	∅
Metadaten													
(interne) Abkürzungen											1	2	8,67
Analysemethode	33	30	12	1	4	3	5	17	22	38	46	216	7,78
(Mess-) Temperatur								2	1	1		7	7,82
Basissatz											1	3	8,75
Bearbeiter							2			2		6	7,6
Simulationsmethode											1	7	8,88
Datum	1							1	2	7	4	21	8,2
Initialien											1	3	8,75
IR- Spektroskopie								1		1	1		6,67
Kinetiken												3	9
Kristallstruktur											1	3	8,75
Laborjournalkürzel	1										1	5	8,83
Literaturzitate												3	9
Lösungsmittel								1		1		3	6,6
Massenspektrometrie										1	1	4	8,5
Datum der Messung										2		2	8
experimentelle Parameter											2	14	8,88
Methode												6	9
NMR												8	9
Parameter		1									1	2	8,67
Probe	1											7	9
Probenbezeichnung	1											16	9
Probennummer										3	2	13	8,56
Probenvorbereitung										1		4	8,6
Projekt									1	1	1	4	8,14
Raumgruppe										1		2	8,33

Bewertung	-9	-1	0	1	2	3	4	5	6	7	8	9	∅
Metadaten													
Reaktionsbedingungen												3	9
Simulationsparameter												7	9
Strukturformel	40	34	51	12	17	21	15	20	12	24	34	147	5,96
Summenformel	43	31	52	19	24	33	25	33	23	30	30	84	5,03
Uhrzeit	1					1					1	3	7,6
Zeitangaben											2	1	8,33

Bei der Berechnung der durchschnittlichen Bewertung wurden die Antworten -9 (nicht beantwortet) und -1 (kann ich nicht beurteilen) außen vorgelassen. In Tabelle 3 zeigt sich, dass die Teilnehmenden durchschnittlich die Metadaten: »Simulationsparameter, Reaktionsbedingungen, Probenbezeichnungen, Probe, Methode, NMR, Literaturzitate und Kinetiken« mit der Topbewertung von 9 am höchsten bewertet haben. Während der allgemeine Begriff »Methode« mit 9 bewertet wurde, wurde der vorgegebene Terminus »Analysemethode« mit 7,78 und die »IR-Spektroskopie« als konkrete Methode mit 6,67 bewertet. Diese Abweichungen können z.B. durch die unterschiedliche Teildisziplinangehörigkeiten der Teilnehmenden zustande kommen. In der einen Teildisziplin zählt das Metadatum mit zu den Hauptkriterien und in einer anderen wird es in einzelnen Fällen betrachtet.

Es zeigt sich, dass die drei am häufigsten genannten Metadaten: »Datum« (36), »Probennummer« (18) und »Probenbezeichnung« (17) zu den administrativen Metadaten gehören. Bei den chemiespezifischen Metadaten stechen die experimentellen Parameter mit 16-facher Nennung heraus. Alle weiteren Metadaten werden weniger als 11-mal genannt.

Bei dieser Frage wurde der Fokus auf den Umfang der Metadaten zur Auffindbarkeit der Daten gelegt. Zusammenfassend lässt sich sagen, dass administrative, Relationship- sowie chemische Metadaten erfasst werden, wobei die drei häufigsten genannten Metadaten »Datum, Probennummer & Probenbezeichnung« zu den Administrativen zählen. Das Verwenden von Synonymen ist mehrfach zu beobachten und die genannten Metadaten weisen eine große Granularität von allgemeinen bis chemie-spezifischen Metadaten auf. Außerdem zeigen die Metadaten eine hohe Heterogenität, die aufgrund der vielen unterschiedlichen Teildisziplinen auftritt. Das EVA-Prinzip kann unterstützen die Metadatenelemente in einem Metadatenschema zu ordnen und somit besser auffindbar und interoperabel machen. Zuletzt ist die stark zurückhaltende Nennung von Metadaten anzuführen, die die Auffindbarkeit sowie nachfolgend die Interoperabilität und Nachnutzbarkeit signifikant erschwert.

7.5 Verständlichkeit der Daten und Laborbücher für den Datenzugriff

Aus der FAIR-Perspektive ist das Kriterium A.2 für die Zugänglichkeit von Daten, dass die Metadaten mindestens genauso lange, wenn nicht sogar länger erhalten bleiben als die Daten selbst (vgl. Kapitel 0). Durch den dauerhaften Zugang der Metadaten und insbesondere der chemiespezifischen Metadaten können die Experimente verstanden und reproduziert werden. Die dauerhafte Zugänglichkeit der Daten und Metadaten zu schaffen, ist Aufgabe der Forschenden, die durch fachspezifische Angebote von Infrastruktur-Einrichtungen unterstützt werden. Chemiespezifische Angebote wie Laborinformationsmanagementsysteme (LIMS) oder ELNs befinden sich derzeit noch im Aufbau oder in der Weiterentwicklung (vgl. zweite Frage dieses Blocks).

Wie zuvor erwähnt, haben die Forschenden auch für die Verständlichkeit der Daten zu sorgen. Aus FAIR- und Infrastruktur-Perspektive werden Daten über ihre Metadaten verständlich, allerdings liegt die Verständlichkeit von Daten im Auge des Betrachters. In der vorangegangenen Frage wurde explizit nach Metadaten gefragt und 70 % der Teilnehmenden haben maximal zwei Metadaten angegeben, wodurch Daten nicht ausreichend verständlich sein können oder doch? Aus den angegebenen Metadaten und

der Granularität wird nur teilweise deutlich, was alles erfasst wird. Daher wurde für das Verständnis aus Sicht der Teilnehmenden die Frage *Könnte eine Wissenschaftlerin/ ein Wissenschaftler innerhalb oder außerhalb Ihres Instituts mit den bisher von Ihnen mitgelieferten Metadaten, Ihre Daten verstehen?* als Einzelauswahlfrage entwickelt, die von den Teilnehmenden nicht beantwortet werden musste.

Wie in Abbildung 13 erkennbar ist, sagten 149 Teilnehmende (35 %), dass ihre Daten innerhalb ihres Institutes, ihrer Arbeitsgruppe, ... ohne weitere Metadaten verstanden werden. Weitere 120 Teilnehmende (29 %) geben an, dass ihre Daten sowohl innerhalb als auch außerhalb ihres Institutes, ihrer Arbeitsgruppe, ... zu verstehen sind.



Abbildung 13: Auswertung der Frage – „Könnte eine Wissenschaftlerin/ ein Wissenschaftler innerhalb oder außerhalb ihres Instituts mit den bisher von Ihnen mitgelieferten Metadaten, Ihre Daten verstehen?“; Antwortmöglichkeiten & relative Häufigkeit der Teilnehmenden

Lediglich eine Minderheit von 38 Teilnehmenden (9 %) beantworteten die Frage mit Nein. Eine Korrelation mit der Position der Teilnehmenden zeigte, dass 3 Professorinnen und Professoren, 29 wissenschaftliche und 2 nicht-wissenschaftliche Mitarbeitende sowie 4 »Andere« nicht überzeugt sind, dass ihre Daten, so wie sie vorliegen, durch Dritte verstanden werden. Die Unsicherheit am Verständnis ihrer Daten ist im Verhältnis bei den wissenschaftlichen Mitarbeitenden am größten (vgl. Tabelle 4).

Tabelle 4: Korrelation: Verständlichkeit der Daten zu Position der Teilnehmenden

	Professorinnen und Professoren	Wissenschaftliche Mitarbeitende	Nicht- wissenschaftliche Mitarbeitende	Andere
An der Umfrage teilgenommen	1	7	1	0,2
Unverständlich- keit der Daten	1	10	0,6	1

Festzuhalten ist, dass 64 % aller Teilnehmenden ihre Daten zumindestens innerhalb einer geschlossenen Gruppe als verständlich empfinden, welches im Widerspruch zu der maximalen Angabe von zwei Metadaten bei 70 % der Teilnehmenden steht.

Um Forschungsdaten FAIR zugänglich zu machen, sollen sie über Standard-Kommunikationsprotokolle heruntergeladen und lokal verwendet werden können. Allgemeine Umfragen wie die österreichweite Befragung zu *Forschende und ihre Daten*¹⁰², an denen Chemikerinnen und Chemiker teilgenommen haben, zeigen, dass es in der Chemie hauptsächlich keinen Zugriff auf Forschungsdaten gibt. Generell zeigt sich in der Umfrage, dass ein großer Teil der Datenweitergabe (54 %) per Mail oder physischem Datenträger erfolgt. In der Umfrage zu *Forschungsdatenspeicherung – Praxis und Bedarfe* vom UNEKE-Projekt¹⁰³ zeigt sich, dass ein Zugriff auf die meisten Daten in einer geschlossenen Gruppe erfolgt und dieser Zugriff nicht kontrolliert wird.

Eine Möglichkeit in der Chemie Daten innerhalb und außerhalb einer geschlossenen Umgebung zu teilen, bietet ein elektronisches Laborbuch (ELN). Laborbücher werden in der Chemie klassischer Weise zur Dokumentation der Experimente, Messungen, Ergebnisse, ... verwendet. Die zweite Frage *Was für eine Art Laborbuch (Papier oder elektronisches Laborbuch) verwenden Sie? Welche Gründe sprechen für die jeweilige Nutzung?* soll ermitteln, wie weit die Nutzung eines Papier- oder elektronischen Laborbuchs verbreitet ist und warum das Jeweilige bevorzugt wird.

¹⁰² Bauer et al. 2015.

¹⁰³ Brenger et al. 2019.

Die Frage ist als Einfachauswahl-Frage mit Drop-Down-Auswahl und einem Lückentext konzipiert. Ein Überspringen der Frage war möglich.

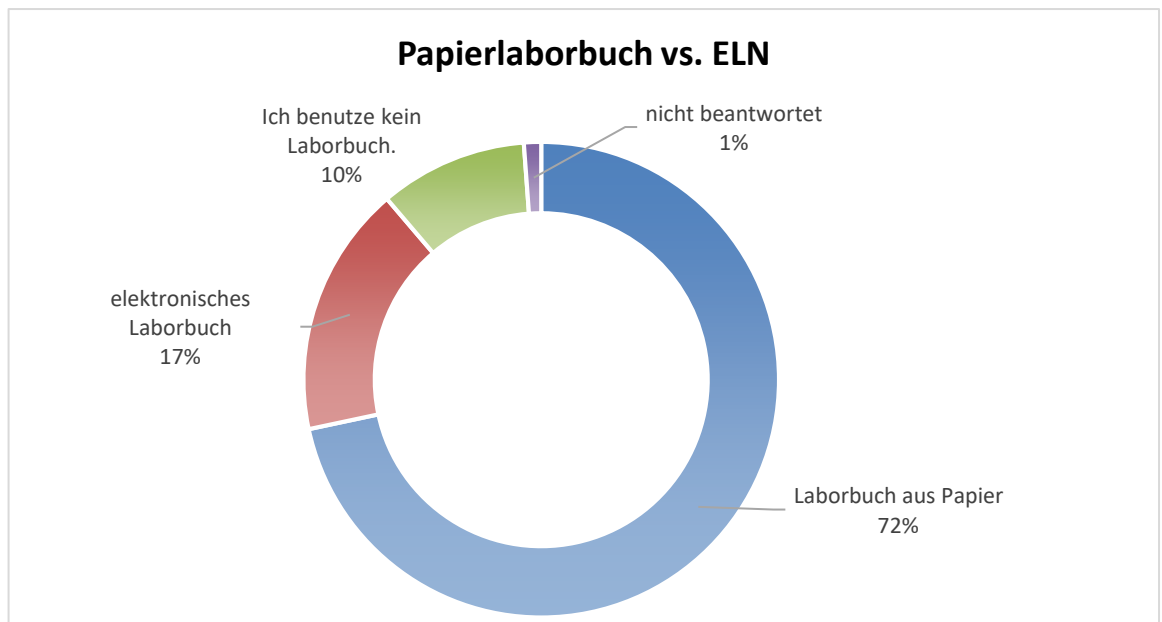


Abbildung 14: Auswertung zur Frage - „Was für eine Art Laborbuch (Papier oder elektronisches Laborbuch) verwenden Sie?“

Eine deutliche Mehrheit aller Teilnehmenden (306; 72 %) verwenden ein Laborbuch aus Papier (vgl. Abbildung 14). Knapp jede 6. Teilnehmende/ jeder 6. Teilnehmender (73; 17 %) arbeitet mit einem ELN und 43 Teilnehmende (10 %) benutzen kein Laborbuch. Im nächsten Schritt wurden die Teilnehmenden gebeten folgenden Satz zu vervollständigen: *Ich verwende es, weil* Für ein Laborbuch in Papierformat wurden 340 Gründe (vgl. Abbildung 15) und für das Arbeiten mit einem ELN 103 Gründe (vgl. Abbildung 16) aufgezählt. 10 Gründe wurden gegen die Verwendung eines Laborbuchs angeführt (vgl. Abbildung 17). Bei den Abbildung 15 bis Abbildung 17 sind die Begründungen aufgeführt, die mindestens von 2 % der Teilnehmenden der jeweiligen Gruppe¹⁰⁴ genannt wurden. Alle weiteren Begründungen wurden aufgrund der Nennung $\leq 2\%$ zu »Andere« zusammengefasst.

¹⁰⁴ Gruppen = Papierlaborbuch, ELN, kein Laborbuch

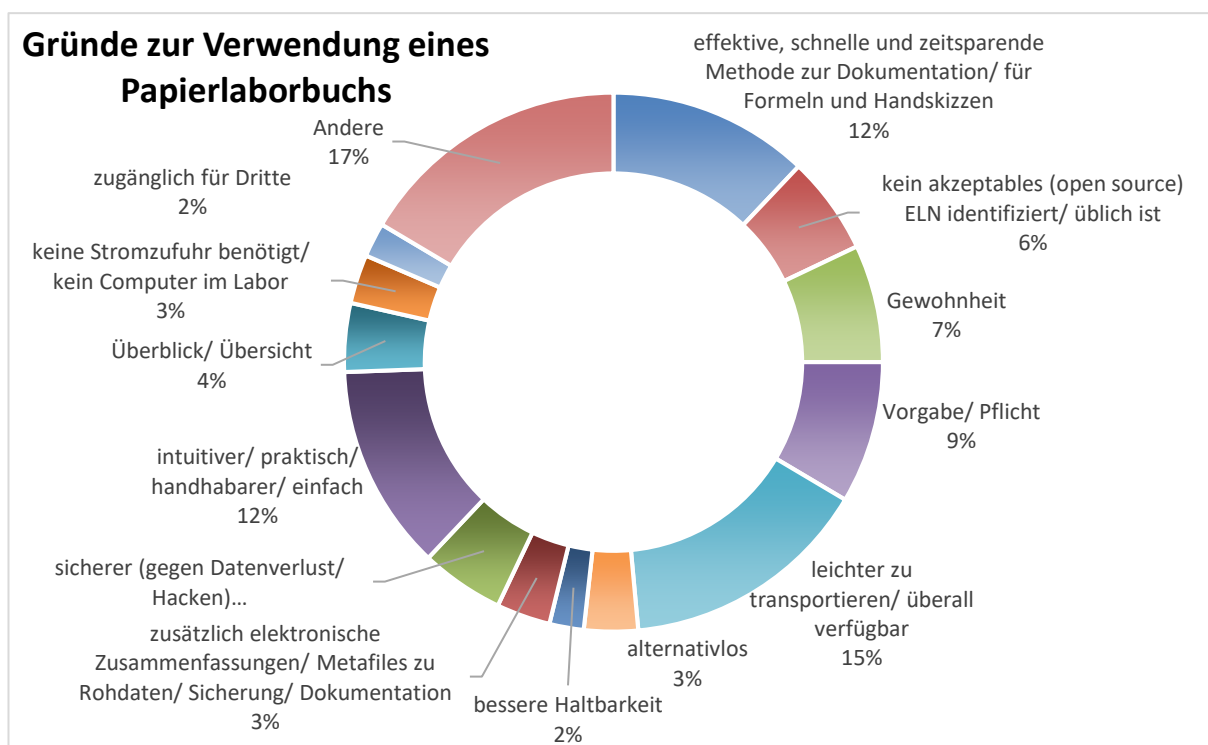


Abbildung 15: Gründe zur Verwendung eines Laborbuchs in Papierformat

Im Vergleich der Argumentation für ein Laborbuch in Papierform oder eines ELNs zeigt sich, dass auf der Seite des Papierlaborbuchs die Gründe »überall verfügbar« (15 %), »effektive, schnelle und zeitsparende Methode« (12 %) sowie die »intuitive, praktische und einfache Handhabung« (12 %) am häufigsten vertreten sind. Ähnlich sieht es beim ELN aus, wo die effektive, schnelle und zeitsparende Methode (10 %) am meisten genannt wird, gefolgt von der besseren »Auffindbarkeit/ Durchsuchbarkeit« (9 %) und dem Komfort, dass es »überall verfügbar« (8 %) ist. Die »einfache und praktische Handhabung« (7 %) wurde am 4. häufigsten genannt. Es zeigt sich, dass meistens die gleichen Argumente sowohl für das Papierlaborbuch sowie für das ELN geliefert wurden. Teilweise wurden Zusätze hinter die einzelnen Gründe geschrieben, die beim Clustern verallgemeinert wurden. Aus diesen Zusätzen lässt sich schließen, dass z.B. der Grund »überall verfügbar« in bestimmten Teildisziplinen wie der Organischen Chemie (OC) und der Anorganischen Chemie (AC) stärker ins Gewicht fallen als in der Physikalischen Chemie (PC). Beim nasschemischen Arbeiten wie es in der OC und AC überwiegend der Fall ist, sind seltener Computer im Labor vorhanden, da diese z.B. durch das Tragen von Handschuhen oder durch das Verschütten einzelner Substanzen leicht kontaminiert oder sogar beschädigt werden können. In der PC hingegen sind viele Labore mit Computern ausgestattet, da die einzelnen Untersuchungsmethoden per Computer

angesteuert werden. Die Handhabung und die Verfügbarkeit eines ELNs ist hier deutlich komfortabler.

Bei den Gründen zur Verwendung eines Papierlaborbuchs findet sich auch der Grund »kein akzeptables (Open Source) ELN identifiziert«. Dieses weist darauf hin, dass die Vorteile eines ELNs wie z.B. die »Verknüpfung der Daten untereinander« (6 %) oder die »Durchsuchbarkeit der Daten« (9 %), erkannt wurden. Beispiele für ELNs sind *Labfolder*, *CERF*, *eLabFTW*, *SciNote*¹⁰⁵ oder *Chemotion*¹⁰⁶. Die Weiterentwicklung einzelner ELNs mit dem Fokus der Anpassung auf die unterschiedlichen Bedürfnisse der Forschenden der einzelnen Teildisziplinen und die Überprüfung der FAIRness einzelner ELNs werden Aufgaben der Initiative NFDI4Chem sein¹⁰⁷. Weitere Aufgaben werden das Schaffen von Awareness für ELNs sein, da die Umfrage zeigt, dass 7 % der Antworten »Gewohnheit« und die dezidierte »Vorgabe/ Pflicht« (9 %) zur Nutzung eines Papierlaborbuchs waren. Als Beispiel einer Awarenessmaßnahme und Hilfestellung sei der *ELN-Wegweiser*¹⁰⁸ für die Lebenswissenschaften der *ZBMed* zu nennen.

¹⁰⁵ Wikipedia 2019b.

¹⁰⁶ Karlsruhe Institute of Technology, Institute for Organic Chemistry und German Research Foundation o.J.

¹⁰⁷ Initiative NFDI4Chem 2019.

¹⁰⁸ ZB MED - Informationszentrum Lebenswissenschaften et al. 2019.

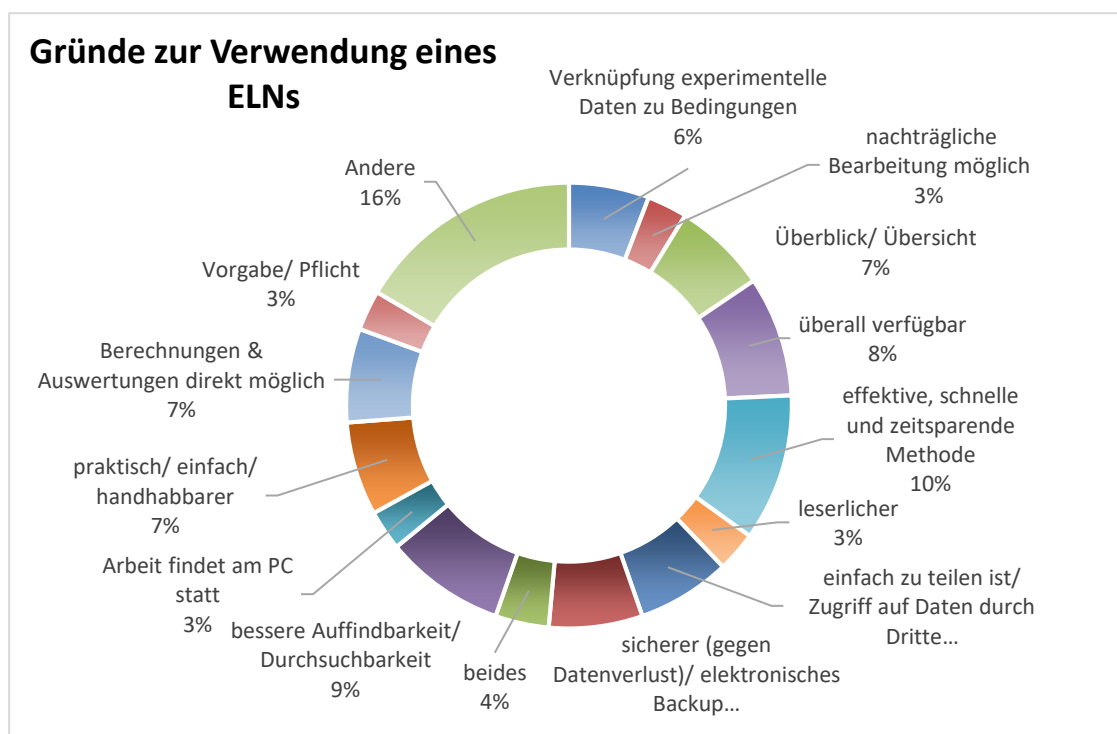


Abbildung 16: Gründe zur Verwendung eines elektronischen Laborbuchs

Von Bedeutung bei den Gründen für ein Papierlaborjournal oder ein ELN ist die Betrachtung der Umgebungsbedingungen.

Gleiches spiegelt sich auch bei den Gründen zur Nutzung keines Laborbuchs wider. Hier wird die nicht vorhandene Labor-Arbeit (27 %), die direkte Weiterverarbeitung der Daten (18 %) und die Verknüpfung der

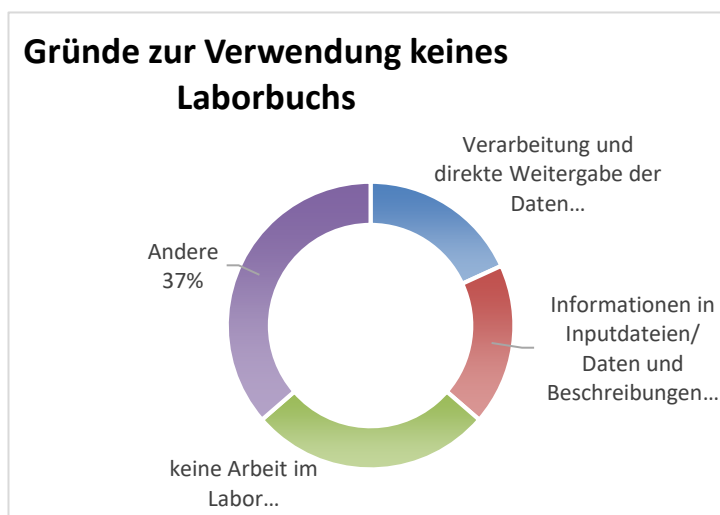


Abbildung 17: Gründe zur Verwendung keines Laborbuchs

Inputdateien mit Beschreibungen (18 %) erwähnt, wodurch die Verwendung eines Laborbuchs überflüssig erscheint.

Zusammenfassend lässt sich sagen, dass die Mehrzahl der Teilnehmenden (70 %) ihre Daten zumindestens innerhalb einer geschlossenen Gruppe als verständlich für Dritte einstufen, so dass, wenn der Zugang zu den Daten gewährt wird, die Daten nutzbar sind. Ein (permanenter) Zugang zu den Daten kann in der Chemie z.B. durch ein ELN generiert werden. Einige der ELNs bieten API-Schnittstellen an, wodurch eine Einbindung von z.B. ePICs möglich wäre und ein permanenter Zugang zu den Daten und Metadaten geschaffen werden könnte. Wie sich zeigte, sind ELNs zu 17 % verbreitet. Klassischerweise kommt bei den meisten Teilnehmenden (72 %) im Labor das Papierlaborbuch zum Einsatz. 6 % der Teilnehmenden haben gezeigt, dass sie Interesse an ELNs haben, nur noch kein für sie brauchbares gefunden haben. Ein in der Chemie alltägliches Tool FAIR nutzbar zu gestalten und diese zu verbreiten, stellt eine Aufgabe der Initiative NFDI4Chem dar.

7.6 Strukturierung der Daten und Vorgaben

Die Strukturierung oder die Benennung der Daten gibt weiteren Aufschluss über die Metadaten sowie über das Arbeiten in der Chemie. Erfolgt eine Verlinkung von Daten untereinander oder werden Daten einzeln erfasst?

Dieser Block besteht aus zwei aufeinander aufbauenden Fragen. Die erste Frage *Neben*

den Metadaten ist auch die Strukturierung ihrer Daten ein wichtiges Element.

Wie strukturieren bzw. benennen Sie ihre Daten?

bietet eine Mehrfachauswahl mit acht vorgegebenen Metadaten und drei offene Texteingabefelder.

Die Frage musste nicht

durch die Teilnehmenden beantwortet werden. Die meisten Teilnehmenden (75 %)

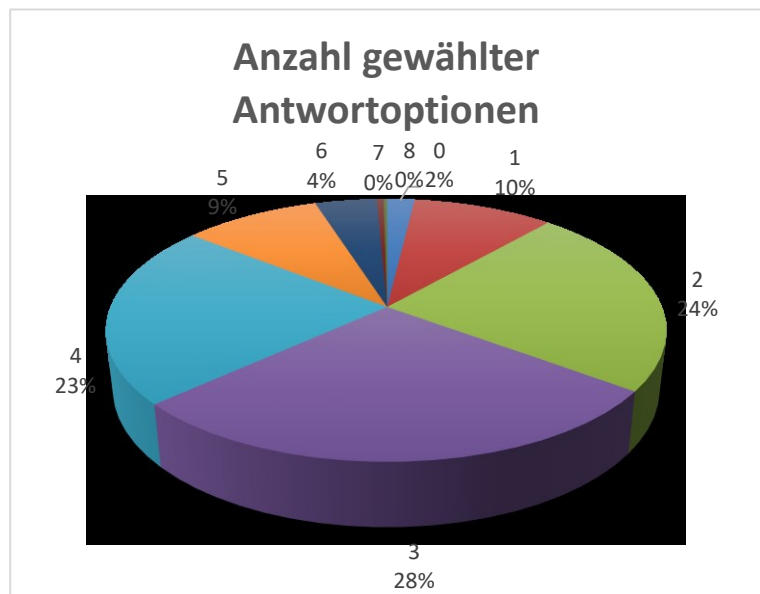


Abbildung 18: Auswertung der Frage – „Wie strukturieren bzw. benennen Sie Ihre Daten?“; Anzahl gewählter Antwortoptionen & relative Häufigkeit der Teilnehmenden

haben mindestens 2 Antwortoptionen gewählt (vgl. Abbildung 18). Damit sind im Durchschnitt 3,08 Metadaten im Dateinamen oder in der Struktur aufgenommen, was im Vergleich knapp einem ganzen Metadatum mehr entspricht als bei der Frage nach der Auffindbarkeit der Daten (durchschnittliche Metadatennennung 2,14).

In der Auswertung der Antwortoptionen zeigt sich, dass insgesamt 1316 Antworten gegeben wurden. Als das wichtigste Metadatum kristallisiert sich eindeutig das »Datum der Erstellung« mit 318-facher Nennung (24 %) raus. Anschließend wurde fast gleichwertig der Name des Forschenden« (229; 17 %), die »Projektnummer« (206; 16 %) und die »Methode der Datenerhebung« (204; 16 %) angegeben. Mit größerem Abstand folgt die »Versionsnummer« (132; 10 %), die »Summenformel nach IUPAC« (73; 6 %) und die »CAS-Nummer« (15; 1 %). Von sehr geringer Bedeutung scheinen die »Nutzungsrechte« zu sein, diese wurde 2-mal (0,15 %) genannt (vgl. Abbildung 19).

Strukturierung/ Benennung der Daten

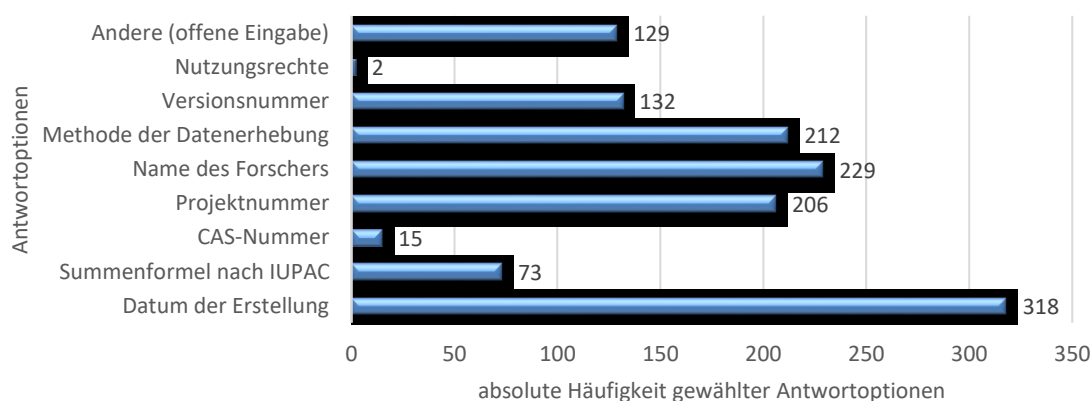


Abbildung 19: Auswertung der Frage I001 – Wie strukturieren bzw. benennen Sie ihre Daten?

Bei der offenen Eingabe wurden insgesamt 137 Metadaten eingetragen, wobei 8-mal die »Analyse-/ Berechnungsmethode« genannt wurde, die dann zur »Methode der Datenerhebung« hinzugezählt wurden. Die vier am häufigsten bei der offenen Eingabe genannten Metadaten sind Tabelle 5 zu entnehmen:

Tabelle 5: am meisten genannte Metadaten bei der offenen Texteingabe

Genanntes Metadatum	Absolute Häufigkeit	Relative Häufigkeit (in %)
Probenbezeichnung	19	1
Probennummer	11	1
Experimentellen Parameter	10	1
Laborbuchnummer	9	1

Die weiteren Metadaten wurden zwischen 1- bis 5- mal gelistet. Zu beobachten ist, dass die am häufigsten genannten Metadaten bei dieser sowie bei der Frage zur Auffindbarkeit der Daten identisch sind. Die Strukturierung nach Datum, Projektnummer, Probennummer oder Laborbuchnummer lässt auf ein chronologisches Arbeiten in der Chemie schließen. Verknüpfungen zwischen einzelnen Datensätzen wie verschiedene Spektren oder Chromatogramme zu einer bestimmten Probe werden nicht erwähnt. Dieses ist lediglich bei der Frage nach den Laborbüchern bei den ELNs angeklungen.

Die zweite Frage in diesem Block: *Strukturieren und/ oder beschreiben Sie Ihre Daten mit Metadaten nach bestimmten Vorgaben, Absprachen, usw.?* zielt darauf ab, ob einheitliche Vorschriften zu Strukturierung der Daten sowie zur Erfassung von Metadaten vorhanden sind.

Die Frage wurde mit Mehrfachauswahl angelegt und konnte übersprungen

werden. Durchschnittlich wurden 1,7 Antworten optiert, dabei haben 177 Teilnehmende (41 %) zwei und 176 Teilnehmende (41 %) eine Antwortoption gewählt (vgl. Abbildung 20).

Anzahl gewählter Antwortoptionen

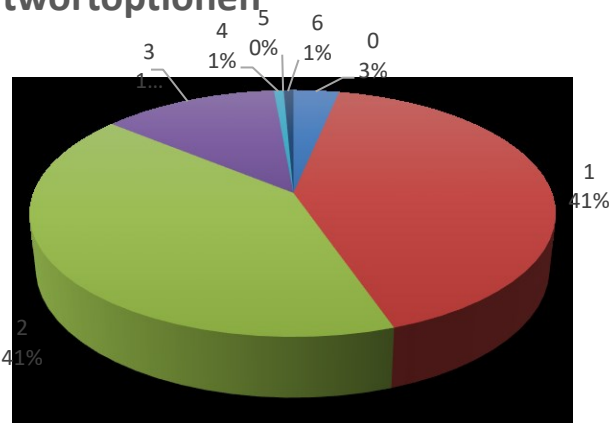


Abbildung 20: Auswertung der Frage - „Strukturieren und/ oder beschreiben Sie Ihre Daten mit Metadaten?“, Anzahl gewählter Antwortoptionen & relativen Häufigkeit der Teilnehmenden

Insgesamt wurden 721 Antworten abgegeben, wobei 368 Antworten (51 %) auf keine Vorgaben zur Strukturierung und Metadatenerfassung entfallen (vgl. Abbildung 21). Bei 40 % gibt es einheitliche Vorgaben zur Strukturierung und Metadatenerfassung, wobei 42-mal (6 %) diese Vorgaben als sinnlos erachtet werden. Weitere 61-mal (9 %) wird einer Absprache unter Kolleginnen und Kollegen getroffen. Die gegebenen Vorgaben und Absprachen ermöglichen eine Interoperabilität der Daten in den jeweiligen geschlossenen Gruppen (Arbeitskreis, Institut, Teildisziplin). Da die Vorgaben überwiegend innerhalb einer Arbeitsgruppe oder eines Instituts gegeben werden und kaum aus der Teildisziplin kommen, zeigt das »Minimum of Information« fehlen bzw. nicht verbreitet sind. Eine Interoperabilität in erweiterter Umgebung kann nicht sichergestellt werden.

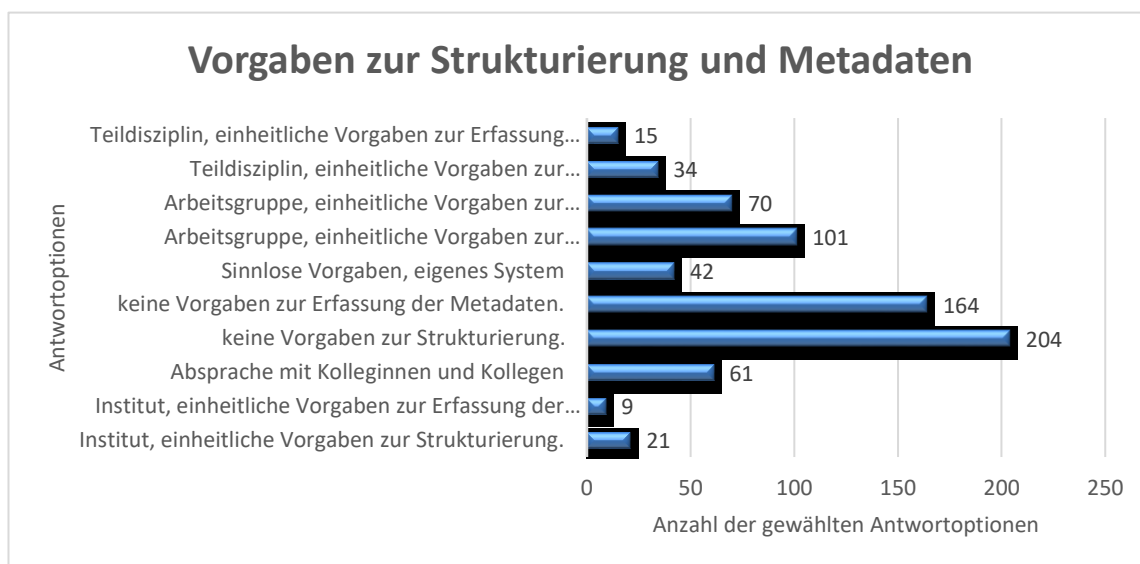


Abbildung 21: Auswertung der Frage - „Strukturieren und/ oder beschreiben Sie Ihre Daten mit Metadaten?“

Zusammenfassend lässt sich sagen, dass das wichtigste Metadatum das Datum der Datenerhebung ist. Dieses sowie die angegebenen unterschiedlichen Nummern (Probennummer, Laborbuchnummer,...) lassen auf eine chronologische Arbeitsweise in der Chemie schließen. Relationen zwischen einzelnen Datensätzen werden nicht erfasst. Weiterhin lässt sich festhalten, dass 40 % der Antworten bei Frage I002 zu Gunsten der einheitlichen Vorgaben zur Strukturierung und Metadatenerfassung ausfielen. Durch diese einheitlichen Vorgaben werden die Daten innerhalb einer geschlossenen Gruppe interoperabel und nachgenutzt, um weitere Forschungsergebnisse zu erzeugen. »Minimum of Information« fehlen oder sind nicht verbreitet.

7.7 Datenweitergabe und Datenpublikation

Die Nachnutzung von Daten und die damit verbundene Datenweitergabe und/ oder Datenpublikation soll in diesem Block beleuchtet werden. Eine Nachnutzung der Daten nach den FAIR-Prinzipien bedeutet unter anderem das Nutzungsrechte für die Daten vergeben werden. Wie sich bereits bei den Fragen zur Auffindbarkeit (vgl. Tabelle 3) sowie zur Strukturierung der Daten (vgl. Abbildung 19) zeigte, wurden »Nutzungsrechte« unter den Metadaten nicht aufgeführt bzw. zweimal als vorgegebene Antwort ausgewählt. Eine Datenweitergabe erfolgt höchstwahrscheinlich ohne eine klare Regulierung der Nutzungsrechte. In der österreichischen Umfrage¹⁰⁹ zeigte sich, dass die meisten Forschenden ihre Daten nur auf Anfrage und an ausgewählte Personen innerhalb ihrer Institution weitergeben. Diese »aktive« Weitergabe der Daten kann aus der Sicht der Forschenden ein ausreichendes Recht auf Nutzung für den Anfragenden implizieren. Bei den verschiedenen befragten Disziplinen war die Verweigerung der Datenweitergabe in der Chemie am höchsten.¹¹⁰ Eine Weitergabe der Daten ist für die chemische Forschung und angrenzende Forschungsbereiche essentiell. Ohne eine Weitergabe der Daten können diese nicht nachgenutzt und keine weiteren neuen Forschungsergebnisse produziert werden.

¹⁰⁹ Bauer et al. 2015.

¹¹⁰ Bauer et al. 2015.

Ebenfalls wurde in der österreichischen Umfrage nach dem Medium der Datenweitergabe gefragt. Durch eine sich weiterentwickelnde Infrastruktur soll mit den Fragen *Wie teilen Sie Daten innerhalb Ihres Instituts?* und *Wie teilen Sie Daten außerhalb Ihres Instituts?* die Aktualität in der Chemie überprüft werden. Wie oben bereits erwähnt, werden die meisten Daten nur innerhalb einer geschlossenen Institution geteilt. Eine Unterscheidung ist sinnvoll, da das Teilen in unterschiedlichen Kreisen verschiedene Möglichkeiten bietet.

Beide Fragen sind gleich aufgebaut und differieren jeweils in einer Antwortoption. Sie sind als Mehrfachantwort mit einem Feld zur offenen Texteingabe angelegt und müssen nicht beantwortet werden.

Auf die Frage *Wie teilen Sie Daten innerhalb Ihres Instituts?* wurden im Durchschnitt 2,4 Antworten gegeben (vgl. Abbildung 22), wobei 227 Teilnehmende (53 %) maximal zwei und zwei Teilnehmende (1 %) sechs Antwortoptionen gewählt haben. Insgesamt wurden 1027 Antwortoptionen in der offenen Texteingabe genannt, daraufhin wurde eine Korrektur der Anzahl der gewählten Antwortoptionen vorgenommen.

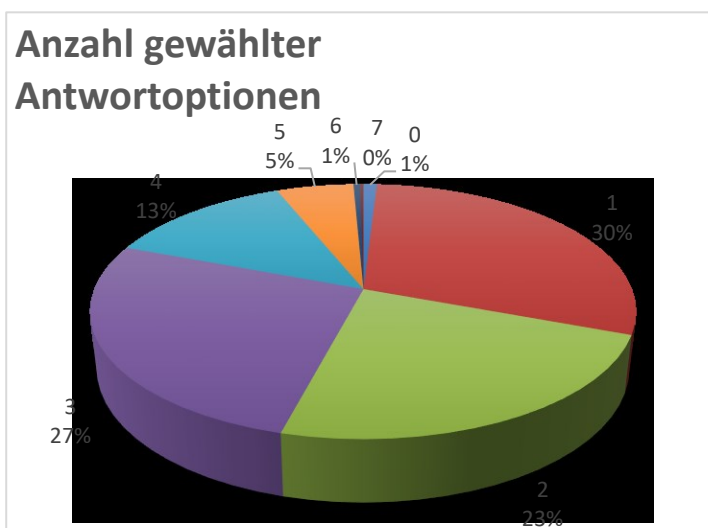
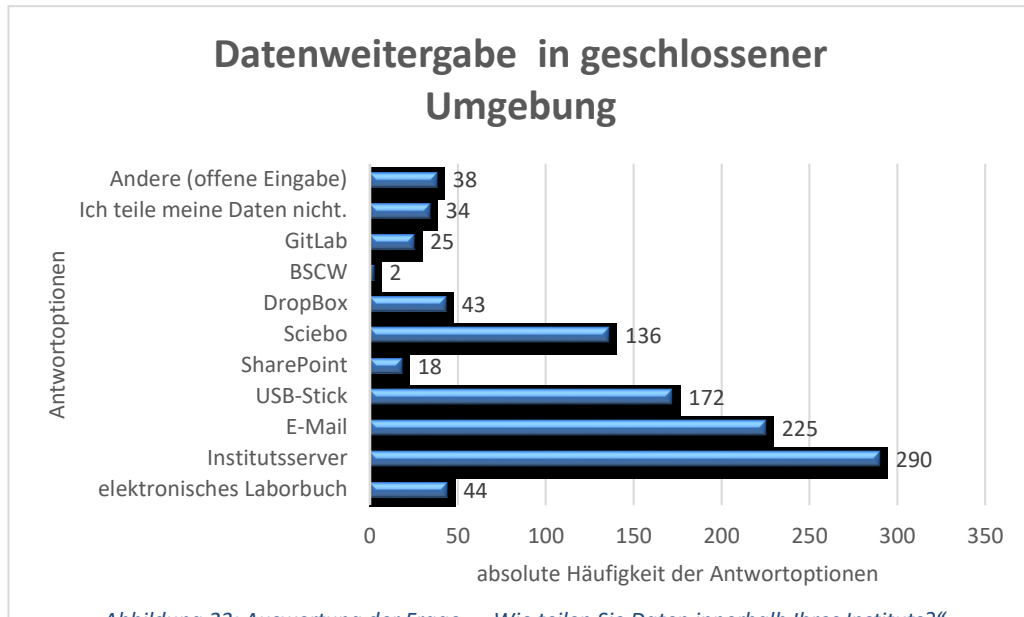


Abbildung 22: Auswertung der Frage – „Wie teilen Sie Daten innerhalb Ihres Instituts?“; Anzahl gewählter Antwortoptionen & relative Häufigkeit



Das Datenteilen mit Kolleginnen und Kollegen des gleichen Instituts erfolgt 290-mal (28 %) über den »Institutsserver« (vgl. Abbildung 23). Am zweithäufigsten werden Daten per »E-Mail« (225; 22 %) weitergegeben, gefolgt von der Weitergabe über einen »USB-Stick« (172; 17 %). Bei Teilen der Daten über kollaborative Plattformen wurde die Campuscloud »Sciebo« 136-mal (13 %) gelistet. Das »elektronische Laborbuch« wurde 44-mal (4 %) genannt, obwohl in Frage nach der Benutzung – Papier oder elektronisch – sich 73 Teilnehmende positiv zu einem ELN geäußert haben. Die Benutzung von »Git« (25; 2 %) ist gegenüber angrenzenden Disziplinen wie z.B. der Physik oder den Ingenieurwissenschaften in der Chemie nicht so stark verbreitet. Unter »Andere« wurden z.B. *Gigamove*, *Slack*, CDs und Ausdrucke genannt, wobei jede Nennung jeweils einmal erfolgte.

Bei der zweiten Frage in diesem Block wurde die Datenweitergabe außerhalb des Instituts betrachtet. Die Frage *Wie teilen Sie Daten außerhalb ihres Instituts?* wurde im Durchschnitt mit 1,5 Antwortoptionen beantwortet. 193 Teilnehmende (45 %) haben nur eine Option und drei Teilnehmende (1 %) haben sechs Antworten gewählt. Insgesamt wurde die Frage mit 787 Antwortoptionen beantwortet, was 23 % weniger Antworten sind

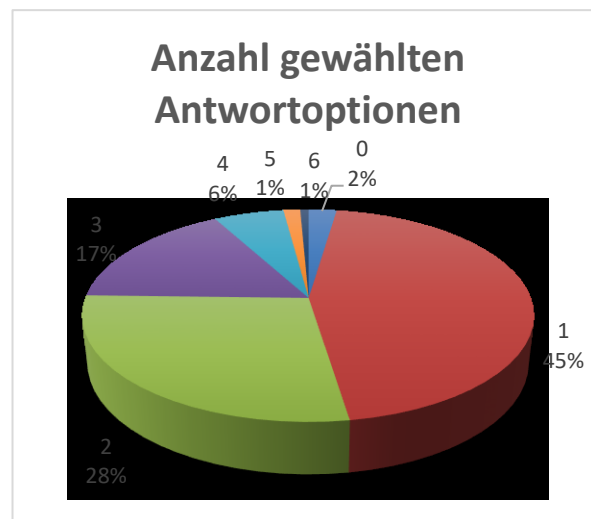


Abbildung 25: Auswertung der Frage – „Wie teilen Sie Daten außerhalb Ihres Instituts?“; Anzahl gewählter Antwortoptionen & relative Häufigkeit der Teilnehmenden

gegenüber der vorangestellten Frage. Eine Korrektur der Antworten wurde vorgenommen, da vorgegebene Antwortoptionen unter »Andere« genannt wurden.

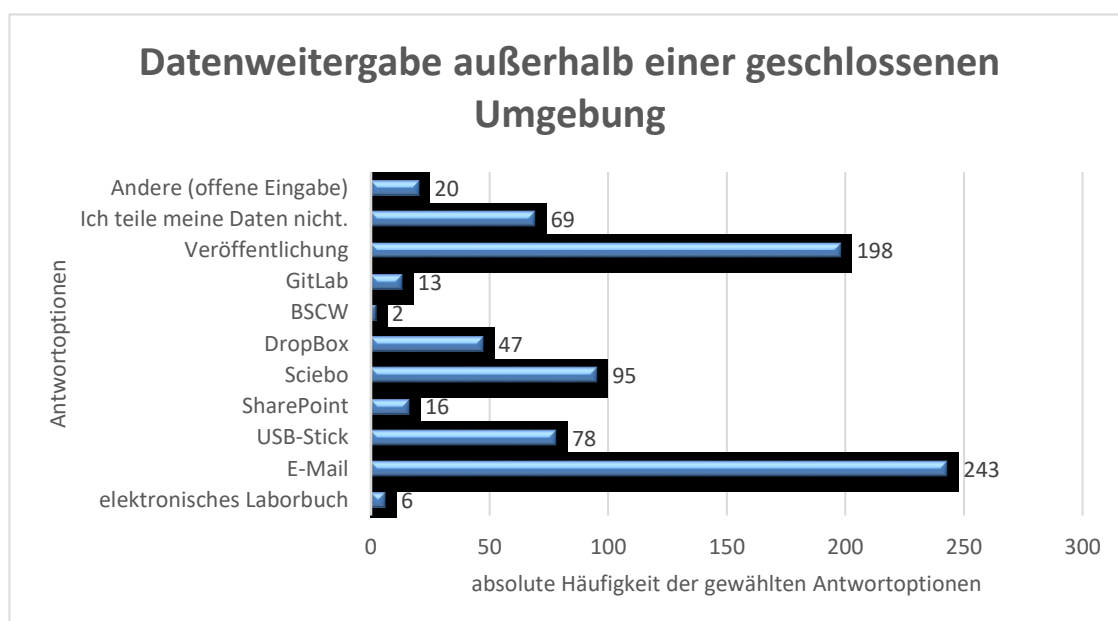


Abbildung 24: Auswertung der Frage – „Wie teilen Sie Daten außerhalb Ihres Instituts?“

Mit einer Mehrheit von 243-mal teilen Chemikerinnen und Chemiker ihre Daten per »E-Mail«, was 30 % entspricht (vgl. Abbildung 24). Gegenüber der generellen österreichischen Umfrage¹¹¹ bei der 54 % der Forschenden Ihre Daten per E-Mail teilen

¹¹¹ Bauer et al. 2015.

und es keine disziplinspezifischen Unterschiede gibt, hat sich diese Weitergabeart in der Chemie verringert.

Nach der Weitergabe der Daten per E-Mail wurde die »Veröffentlichung« (198; 25 %) gelistet. Publierte Daten stehen der Öffentlichkeit zur Verfügung. Dieses untermauert indirekt das Ergebnis, dass unveröffentlichte Daten verschlossen bleiben.¹¹² Ähnliches zeichnet sich auch bei der Antwortoption »Ich teile meine Daten nicht.« ab – bei der externen Datenweitergabe wird diese 69-mal (9 %) und bei der Internen halb so oft, 34-mal (3 %) genannt. Das externe Teilen über kollaborative Plattformen ist verhältnismäßig gleich zum internen Teilen – »Sciebo« (84; 12 %), »Git« (12; 2 %) – das »elektronische Laborbuch« fällt ab auf 6 Antworten (1 %). Unter »Andere« wurde siebenmal *Gigamove* (1 %) sowie z.B. die kristallographischen Datenbanken, mündlich oder die externe Festplatte jeweils einmal aufgeführt.

Die Datenweitergabe unterscheidet sich zur Datenpublikation im Grad der Zugänglichkeit. Bei der Datenweitergabe werden Daten einem geschlossenen Nutzerkreis und bei der Datenpublikation der Öffentlichkeit zugänglich gemacht. Wie sich zeigte, werden Daten in Form von Veröffentlichungen geteilt.

Die Frage *Haben Sie bereits Daten in Form von Rohdaten, bearbeiteten Daten, ... veröffentlicht?* zielt auf das Verhalten sowie auf die Gewohnheit der Forschenden bei der Publikation ab.

Sie wurde als Mehrfach-Frage mit einer offenen Testeingabe als Ausweichoption konzipiert. Die Frage wurde durchschnittlich mit 1,1 Antworten optiert. Sie musste nicht beantwortet werden (vgl. Abbildung 26).

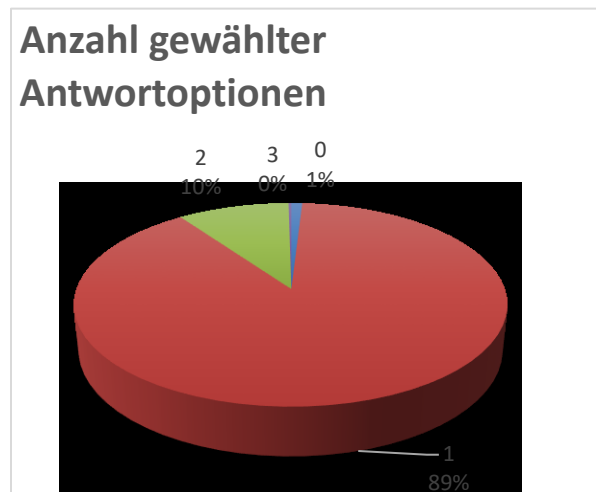


Abbildung 26: Auswertung der Frage – „Haben Sie bereits Daten veröffentlicht?“; Anzahl gewählter Antwortoptionen & relative Häufigkeit der Teilnehmenden

Insgesamt wurden 466 Antworten gegeben, bei denen sich zwei Antwortoptionen rauskristallisierten (vgl. Abbildung 27). Entweder wurden die Daten als

¹¹² Bauer et al. 2015.

»supplement material im pdf-Format zusammen mit einer Textpublikation veröffentlicht« (209; 45 %) oder es wurden bisher »keine Daten veröffentlicht« (194; 42 %). Unter der Option »Andere« wurden als weitere Publikationsformen, die »Publikation als Poster oder Vortrag« sowie als »supplement material im mp4/tif- und maschinenlesbaren Format« genannt.

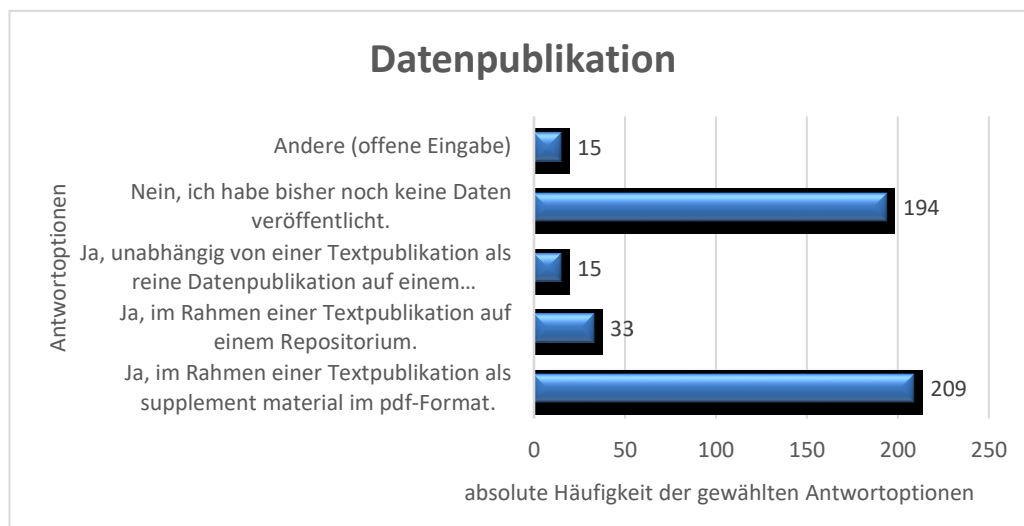


Abbildung 27: Auswertung der Frage – „Haben Sie bereits Daten veröffentlicht?“

Als supplement material oder supporting information zu einer Textpublikation werden z.B. die Vorschriften der Synthese, einige Eigenschaften des Produktes und die Strukturaufklärung veröffentlicht. In Abbildung 28 ist ein $^1\text{H-NMR}$ -Spektrum¹¹³ gezeigt, welches der Strukturaufklärung dient und im Rahmen einer Textpublikation beim *Journal of the American Chemical Society* auf Figshare veröffentlicht wurde. Es zeigt sich, dass ein Bild des Spektrums veröffentlicht ist, was das Spektrum nicht reproduzierbar macht. Im Fließtext sind u.a. die Peaks des $^1\text{H-NMR}$ -Spektrums, das Lösungsmittel und die Pulsfrequenz aufgezählt, der Datensatz zu diesem Spektrum ist nicht zugänglich. Damit sind sowohl das Spektrum wie die weiteren genannten Metadaten nicht maschinenlesbar.

¹¹³ Yan et al. 2019.

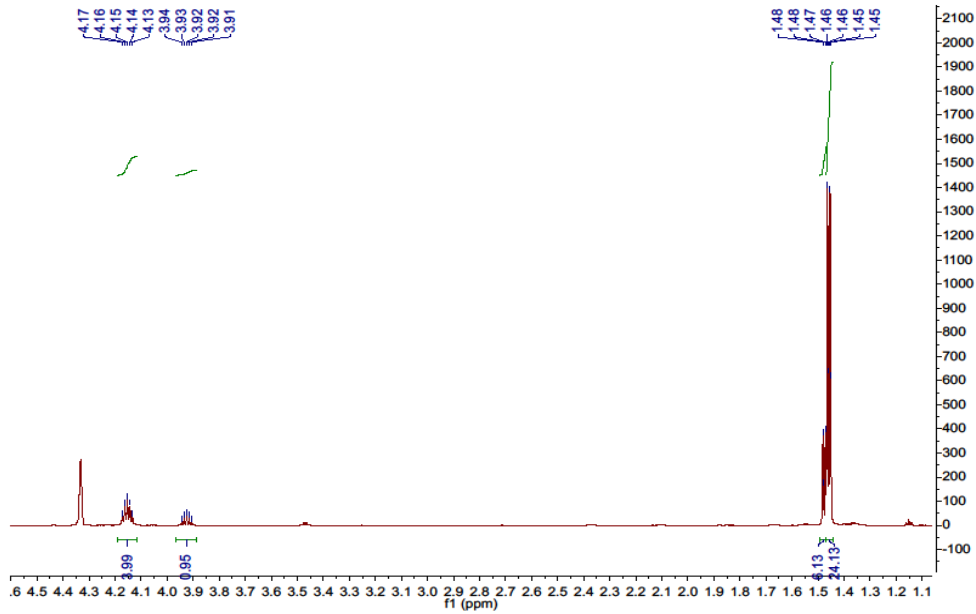


Abbildung 28: Veröffentlichung eines ^1H -NMR-Spektrums von Yichao Yan et al., „Supporting Information for: Mechanism-based design of a high-potential catholyte enables a 3.2 V all-organic nonaqueous redox flow battery,“ Journal of the American Chemical Society, 2019.

Re3data.org¹¹⁴ liefert bei der Suche nach dem Schlagwort »NMR« 12 Resultate für Open Access-Repositorien wie z.B. *NMRshiftDB*¹¹⁵, *chemotion*¹¹⁶ oder die *Spectral Database for Organic Compounds*¹¹⁷, wo die spektralen Datensätze veröffentlicht sind und heruntergeladen werden können. Die Datensätze liegen in offenen Austauschformaten wie JCAMP-DX¹¹⁸ oder NMReDATA¹¹⁹ vor und können nachgenutzt werden.

¹¹⁴ re3data.org Project Consortium o.J.

¹¹⁵ Cologne University et al. o.J.

¹¹⁶ Karlsruhe Institute of Technology, Institute for Organic Chemistry und German Research Foundation o.J.

¹¹⁷ National Metrology Institute of Japan und National Institute of Advanced Industrial Science and Technology o.J.

¹¹⁸ IUPAC o. J.

¹¹⁹ Pupier et al. 2018.

Die letzte Frage *Haben Sie selbst schonmal Daten aus anderen Quellen nachgenutzt?* in diesem Block zielt auf das Nutzungsverhalten der Teilnehmenden von fremden Daten ab.

Sie ist als Mehrfachauswahl-Frage entwickelt. Die Teilnehmenden haben im Durchschnitt 1,9 Antwortoptionen gewählt (vgl. Abbildung 29).

Insgesamt wurden 794 Antworten

gegeben, von denen 83 % positiv für eine Nachnutzung ausfielen (vgl. Abbildung 30). Die meisten Teilnehmenden (256; 32 %) nutzen Daten »innerhalb einer geschlossenen Umgebung« nach, gefolgt von der Datennachnutzung aus einem »supplement material« (198; 25 %). Daten von einem Repository wurden 47-mal (6 %) nachgenutzt, was gegenüber der Datenveröffentlichung (15; 3 %) auf einem Repository dreimal so viel ist.

22-mal (3 %) wurde die Angabe gemacht, dass die Daten nicht ausreichend beschrieben waren und nicht nachgenutzt werden konnten. Dieses unterstützt das Ergebnis der Frage zur Verständlichkeit der Daten, bei dem zu 64 % eine Verständlichkeit der Daten innerhalb einer geschlossenen Gruppe sich rauskristallisierte.

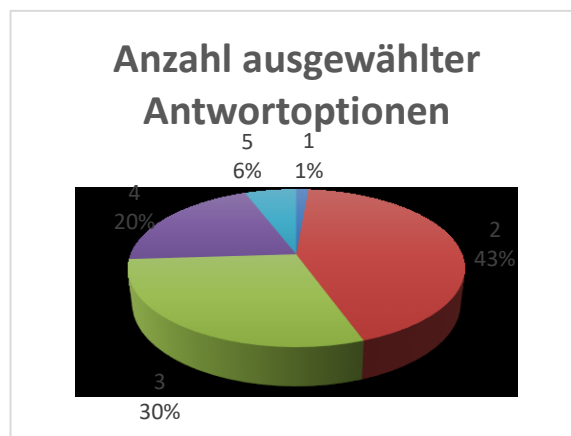


Abbildung 29: Auswertung der Frage – „Haben Sie selbst schonmal Daten aus anderen Quellen nachgenutzt?“; Anzahl gewählter Antwortoptionen & relative Häufigkeit der Teilnehmenden

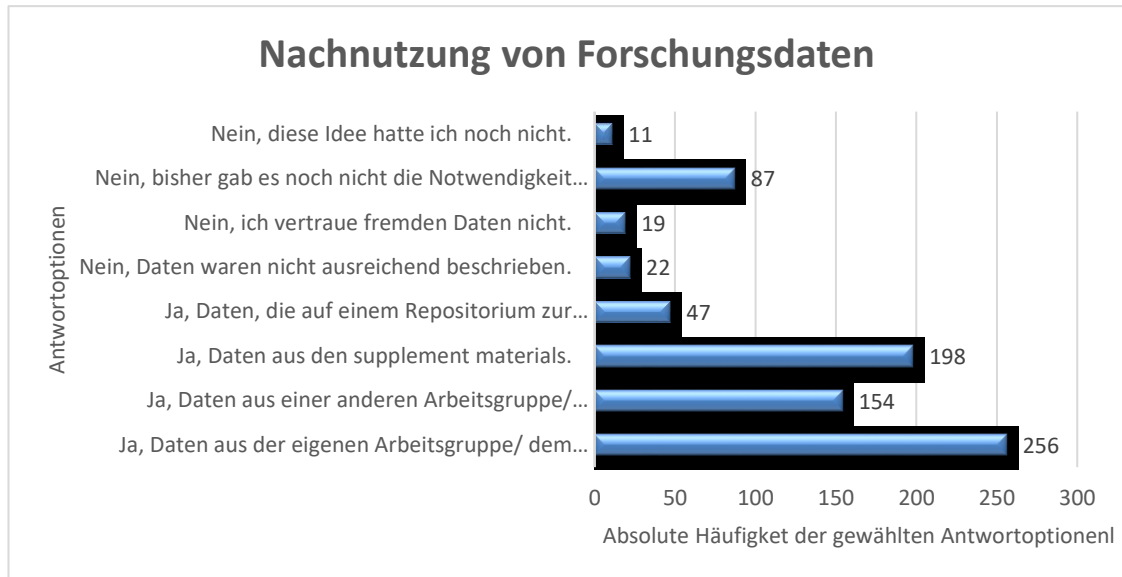


Abbildung 30: Auswertung der Frage - „Haben Sie selbst schonmal Daten aus anderen Quellen nachgenutzt?“

Abschließend lässt sich festhalten, dass eine Datenweitergabe innerhalb einer Arbeitsgruppe/ eines Instituts/ ... am häufigsten (28 %) über den Arbeitsgruppe-/Institutsserver erfolgt. Verlassen Daten die Arbeitsgruppe/ das Institut/ ..., geschieht dieses entweder per E-Mail (30 %) oder in Form einer Veröffentlichung (25 %). Sciebo wird bei beiden Bedingungen verhältnismäßig gleich (12 – 13 %) eingesetzt. Interessanterweise nahm die Beteiligung an den Fragen von interner (1027) nach externer Datenweitergabe (787) ab und sank noch weiter bei der Datenpublikation (466). Daten werden in der Chemie hauptsächlich als supplement material (45 %) veröffentlicht oder sie werden nicht veröffentlicht (42 %). Daten, die als supplement material veröffentlicht werden, können nicht in dem Umfang nachgenutzt werden wie es aus FAIR-Perspektive gewünscht ist, da die Daten z.B. als ausgewertete Bilddateien und nicht als Daten in offenen Formaten vorliegen. 198 Teilnehmende nutzen trotz der Einschränkungen Daten aus Veröffentlichungen nach. Die meisten Daten (32 %) werden intern nachgenutzt.

Es zeigt sich, dass ein Umdenken bei der Publikation erfolgen muss, um chemische Daten nach zu nutzen. Das supplement material sollte nicht aus z.B. Bilder von ausgewerteten Spektren bestehen, sondern sollte die Daten in einem offenen Format oder die persistenten Identifikatoren enthalten.

7.8 Danksagung und Anmerkungen

Zuletzt bestand für die Teilnehmenden die Möglichkeit unter „Für aktuelle Fragen, Bemerkungen und Anregungen finden Sie hier Platz“ eine Rückmeldung geben, wovon 21 Teilnehmende (5 %) Gebrauch gemacht haben.

Die Antworten lassen sich zusammenfassen zu:

- Kritische Betrachtung der Datenveröffentlichung. Das Interesse anderer Forschende an den Daten wird angezweifelt. (2)
- Der Aufwand einer Datenpublikation soll in einem vertretbaren Rahmen liegen. (2)
- Interesse an den Ergebnissen und Daten (1)
- Hinweis auf Cheminformatics. (1)
- Metadaten-Management soll einfach und flexibel sein. (1)
- Ein Biochemiker sowie ein theoretischer Chemiker fanden den Fragebogen nicht teildisziplinspezifisch genug. Die Breite der Chemie wurde zu generell abgedeckt. (2)
- Spezifischere Betrachtung der einzelnen Teildisziplinen wäre wünschenswert und würde das FDM in der Teildisziplin weiterbringen. (2)
- Hochschulweites, open source ELN wäre wünschenswert. (1)
- Individueller Arbeitsstil soll über Vorschriften stehen. (1)
- Weitere Erläuterungen wären wünschenswert gewesen. (1)
- Viel Erfolg. (3)

8. Zusammenfassung und Ausblick

Experimente bilden das Herz der Chemie, entsprechend werden am meisten experimentelle Datensätze erzeugt. Forschungsdaten reichen in der Chemie z.B. von experimentellen über analytische bis zu simulierten Daten und weisen eine hohe Heterogenität auf. Ebenso verhält es sich mit den Metadaten. Bei der Nennung von Metadaten zur Auffindbarkeit und Strukturierung wurden viele Metadaten nur ein- bis dreimal genannt, was auf keine bis willkürliche Strukturierung der Daten schließen lässt. Vorgaben zur Strukturierung gibt es bei 2/5 aller Teilnehmenden, wobei diese Vorgaben

hauptsächlich von der Arbeitsgruppe oder dem Institut gemacht werden. Es mangelt an einem Minimum of Information bzw. an der Verbreitung.

Metadaten schemata sind in der Chemie nicht verbreitet. Das vorgestellte EVA-Prinzip kann die Teilnehmenden unterstützen, ein Bewusstsein für chemiespezifische Metadaten zu entwickeln und bei dem Schritt Metadaten schemata für ein konkretes Experiment, für einen bestimmten Workflow, etc. zu erzeugen. Trotz des Potpourris an Metadaten und den geringen Überschneidungen der Metadaten gehen die meisten Teilnehmenden davon aus, dass ihre Daten innerhalb ihrer Arbeitsgruppe, ihres Instituts, etc. ohne weitere Angaben verständlich sind. Das am häufigsten verwendete Metadatum ist das »Datum«, das von »Probennummer, Probenbezeichnung & Laborbuchnummer« gefolgt wird. Daraus lässt sich ableiten, dass in der Chemie chronologisch gearbeitet wird und Daten untereinander wenig verknüpft sind. Das Abbilden von Relationen zwischen Daten wurde als ein Vorteil eines ELNs aufgeführt. ELNs sind in den chemischen Instituten an nordrhein-westfälischen Hochschulen und außeruniversitären Einrichtungen lediglich zu 17 % verbreitet. Die klassische Arbeitsweise mit einem Laborbuch aus Papier dominiert. Die genannten Gründe für das jeweilige Laborbuch wie überall verfügbar, leichte Handhabbarkeit, etc. waren auf beiden Seiten die gleichen und interessanterweise ähnlich stark vertreten. ELNs können durch die Einbindung von z.B. ePICs¹²⁰ eine dauerhafte Referenzierbarkeit der Daten sowie einen permanenten Zugang zu den Daten und Metadaten schaffen. Der Zugang zu Daten erfolgt verstärkt innerhalb einer Arbeitsgruppe oder eines Instituts. Das Teilen erfolgt überwiegend über Arbeitsgruppen-/Institutsserver. Außerhalb der Arbeitsgruppe, des Instituts, etc. werden Daten vermehrt noch per E-Mail geteilt und häufig in Form einer Veröffentlichung. Viele Daten werden in der Chemie als supplement materials im pdf-Format veröffentlicht, so dass es nicht durch Maschinen ausgewertet werden kann. Dabei gibt es sehr erfolgreiche Beispiele wie die Kristallographie, die seit über 50 Jahren ihre Daten öffentlich auf einem Repository speichern. Insgesamt sind die Teilnehmenden im Punkt Publikation sehr zurückhaltend, welches sich auch in der geringen Beteiligung bei der Frage zur Publikation widerspiegelt (vgl. Abbildung 31). Die Frage hat die durchschnittlich geringste Beteiligung von allen Fragen mit

¹²⁰ ePIC - Persistent Identifier for Research o.J.

Mehrfachauswahl. Obwohl die Daten vermehrt im pdf-Format vorliegen, werden sie durch knapp die Hälfte aller Teilnehmenden nachgenutzt. Chemikerinnen und Chemiker nutzen deutlich häufiger Daten nach als die ihren zu teilen.

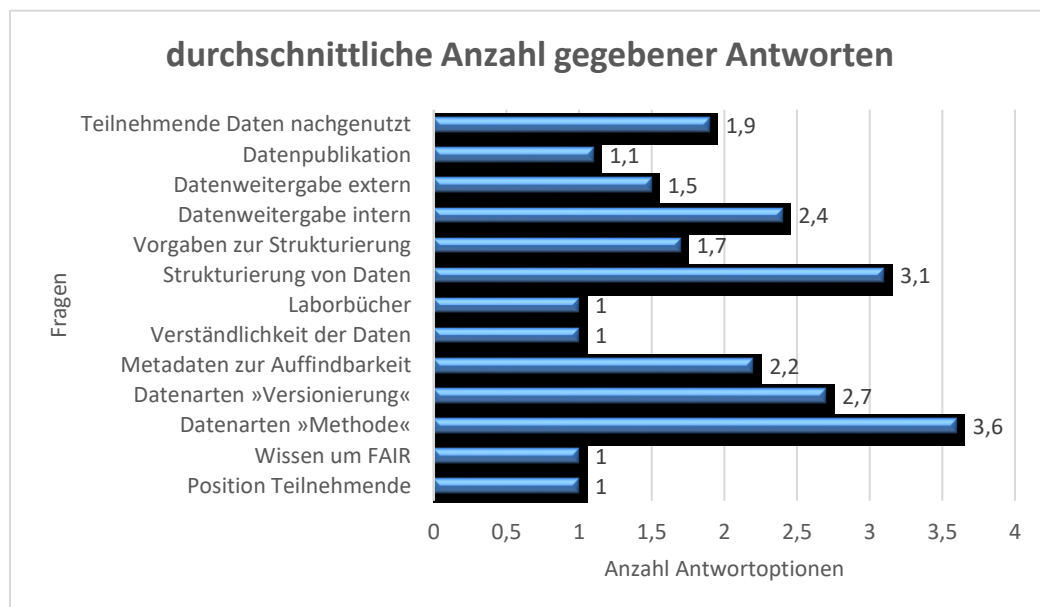


Abbildung 31: durchschnittlich gegebene Antworten

Abbildung 31 verdeutlicht, dass die Teilnehmenden die meisten Antworten bei der Frage nach den Datenarten »Methode«, der Strukturierung sowie bei den Datenarten »Versionierung« gegeben haben. Die Inhalte der Fragen sind den Teilnehmenden durch ihre alltägliche Arbeit am vertrautesten.

Schlussfolgern lässt sich, dass Daten größtenteils innerhalb einer Arbeitsgruppe, eines Instituts, etc. auffindbar, zugänglich, interoperabel und nachnutzbar sind. Allerdings nicht so wie es die FAIR-Prinzipien vorgeben. Die FAIR-Prinzipien hingegen sind 88 % der Teilnehmenden nicht bekannt. Der Fachbereich Chemie hat in NRW, deutschland-, europa- und weltweit einen starken Aufholbedarf auf dem Gebiet FDM, wie es auch *Howes* in ihrem Artikel beschreibt „getting there will be a löng road.“¹²¹

Um diese Unwissenheit mit Wissen und Leben zu füllen, bedarf es neuer Strukturen. Strukturen wie z.B. der NFDI mit der Initiative NFDI4Chem. Aber auch Fachreferentinnen und Fachreferenten der Chemie sind gefragt, diese Lücke mit entsprechenden Awarenessmaßnahmen, Beratungen und Trainings zu schließen oder auch

¹²¹ Howes 2019.

Unterstützung zu leisten beim Aufbau neuer interner Strukturen wie einem Data Stewardship-Programm.

9. Literaturverzeichnis

Allianz der deutschen Wissenschaftsorganisationen (2010): Grundsätze zum Umgang mit Forschungsdaten. DOI: 10.2312/ALLIANZOA.019.

Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V.; Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e.V.; Berufsverband Deutscher Markt- und Sozialforscher e.V. (1999): Standards zur Qualitätssicherung in der Markt- und Sozialforschung, zuletzt geprüft am 27.08.2019.

Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V.; Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e.V.; Berufsverband Deutscher Markt- und Sozialforscher e.V.; Deutsche Gesellschaft für Online-Forschung e.V. (o.J.): Standards zur Qualitätssicherung für Online-Befragungen, S. 1–8, zuletzt geprüft am 26.08.2019.

Bauer, Bruno; Ferus, Andreas; Gorraiz, Juan; Gründhammer, Veronika; Gumpenberger, Christian; Maly, Nikolaus et al. (2015): Forschende und ihre Daten. Ergebnisse einer österreichweiten Befragung - Report 2015. 1.2. Aufl. Koordinationsbüro, e-Infrastructures Austria. Wien, zuletzt geprüft am 08.09.2019.

Baur, Nina; Blasius, Jörg (2014): Methoden der empirischen Sozialforschung. In: Nina Baur und Jörg Blasius (Hg.): Handbuch Methoden der empirischen Sozialforschung, Bd. 48. Wiesbaden: Springer VS, S. 41–62, zuletzt geprüft am 27.08.2019.

Beywl, Wolfgang; Schepp-Winter, Ellen (2000): Zielgeführte Evaluation von Programmen: ein Leitfaden (QS - Materialien zur Qualitätssicherung in der Kinder- und Jugendhilfe., 29).

Brand, Ortrun; Stille, Wolfgang; Schachtner, Joachim (2018): HeFDI – Die landesweite Initiative zum Aufbau von Forschungsdateninfrastrukturen in Hessen. 14-27 Seiten / o-bib. Das offene Bibliotheksjournal / herausgegeben vom VDB, Bd. 5, Nr. 2 (2018). DOI: 10.5282/O-BIB/2018H2S14-27.

Brenger, Bela; Rehwald, Stephanie; Wilms, Konstantin L.; López, Ania; Stieglitz, Stefan (2019): UNEKE: Forschungsdatenspeicherung - Praxis und Bedarfe: Online-Survey 2019. University of Duisburg-Essen, Germany. DuEPublico: Duisburg-Essen Publications online, zuletzt geprüft am 08.09.2019.

Bruno, Ian J.; Groom, Colin R. (2014): A crystallographic perspective on sharing data and knowledge. In: *Journal of computer-aided molecular design* 28 (10), S. 1015–1022. DOI: 10.1007/s10822-014-9780-9.

Coles, Simon (Hg.) (o.J.): Active GO FAIR Implementation Network. Online verfügbar unter <https://www.go-fair.org/implementation-networks/overview/chemistryin/>, zuletzt aktualisiert am 13.09.2019.

Cologne University, Bioinformatics Facility; European Molecular Biology Laboratory, European Bioinformatics Institute; Max-Planck-Institute for Chemical Ecology; German Research Foundation (Hg.) (o.J.): NMRshiftDB. Online verfügbar unter <http://nmrshiftdb.nmr.uni-koeln.de>, zuletzt geprüft am 11.09.2019.

Daudrich; Anna (2018): Umgang mit Digitalen Forschungsdaten in den Geistes- und Sozialwissenschaften. Online verfügbar unter <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=4&ved=2ahUKEwjZus2Yjs7kAhWO-KQKHeh9BOUQFjADegQIARAC&url=https%3A%2F%2Fwww.fdm-bayern.org%2Ffiles%2F2018%2F11%2Fforschungsdatenmanagement-in-den-geisteswissenschaften-an-der-fau-umfrage.pdf&usg=AOvVaw0nrxtJk7FdpS8EAZ02AYUY>, zuletzt geprüft am 13.09.2019.

Deutsche Forschungsgemeinschaft (Hg.) (2019): Absichtserklärungen zum Aufbau einer NFDI. Online verfügbar unter <https://www.dfg.de/foerderung/programme/nfdi/absichtserklaerungen/index.html>, zuletzt geprüft am 12.09.2019.

Digital Curation Center (Hg.) (2004-2019): DCC Curation Lifecycle Model. Online verfügbar unter <http://www.dcc.ac.uk/resources/curation-lifecycle-model>, zuletzt geprüft am 13.09.2019.

Digital Curation Center (Hg.) (2014): Five steps to decide what data to keep. a checklist for appraising research data. Online verfügbar unter <http://www.dcc.ac.uk/resources/how-guides/five-steps-decide-what-data-keep>, zuletzt geprüft am 14.09.2019.

Duden (o.J.): Experiment. Online verfügbar unter <https://www.duden.de/rechtschreibung/Experiment>, zuletzt geprüft am 01.09.2019.

ePIC - Persistent Identifier for Research (o.J.). Online verfügbar unter <https://www.pidconsortium.eu/>, zuletzt geprüft am 09.09.2019.

EUDAT - European Data Project (o.J.). Online verfügbar unter <https://eudat.eu/>, zuletzt geprüft am 15.09.2019.

European Open Science Cloud (Hg.) (2018): European Open Science Cloud. Online verfügbar unter <https://www.eosc-portal.eu/>, zuletzt geprüft am 15.09.2019.

forschungsdaten.info (Hg.) (o.J.): Metadaten und Metadatenstandards. Online verfügbar unter <https://www.forschungsdaten.info/themen/aufbereiten-und-veroeffentlichen/metadaten-und-metadatenstandards/>, zuletzt geprüft am 07.09.2019.

forschungsdaten.org (Hg.) (o.J.): Forschungsdaten. Online verfügbar unter <https://www.forschungsdaten.org/index.php?title=Forschungsdaten&oldid=2482>, zuletzt geprüft am 11.09.2019.

Fühles-Ubach, Simone (2013a): Online-Befragungen. In: Konrad Umlauf, Simone Fühles-Ubach und Michael Seadle (Hg.): Handbuch Methoden der Bibliotheks- und Informationswissenschaft. Bibliotheks-, Benutzerforschung, Informationsanalyse. Berlin: Walter de Gruyter, S. 114–127, zuletzt geprüft am 26.08.2019.

Fühles-Ubach, Simone (2013b): Quantitative Befragungen. In: Konrad Umlauf, Simone Fühles-Ubach und Michael Seadle (Hg.): Handbuch Methoden der Bibliotheks- und Informationswissenschaft. Bibliotheks-, Benutzerforschung, Informationsanalyse. Berlin: Walter de Gruyter, S. 97–113, zuletzt geprüft am 29.08.2019.

Hausen, Daniela Adele; Trautwein-Bruns, Ute (2019): Erste Schritte im FDM. DOI: 10.18154/RWTH-2019-05139.

Hausen, Daniela Adele; Windeck, Jürgen (2018): Entwicklung eines Blended Learning Kurses zum Forschungsdatenmanagement an der RWTH Aachen University. In: *o-bib* 5 (3). DOI: 10.5282/o-bib/2018H3S17-31.

Helmut Löbner (2019): Online-Befragung. Aachen, 10.04.2019. E-Mail.

Herres-Pawlis, Sonja; Koepler, Oliver; Steinbeck, Christoph (2019): NFDI4Chem: Shaping a Digital and Cultural Change in Chemistry. In: *Angewandte Chemie (International ed. in English)* 58 (32), S. 10766–10768. DOI: 10.1002/anie.201907260.

Hochschulrektorenkonferenz (Hg.) (2014): Management von Forschungsdaten. eine zentrale strategische Herausforderung für Hochschulleitungen. Bonn. Online verfügbar unter <https://www.hrk.de/positionen/beschluss/detail/management-von-forschungsdaten-eine-zentrale-strategische-herausforderung-fuer-hochschulleitungen/>, zuletzt geprüft am 15.09.2019.

Howes, Laura (2019): Chemistry data should be FAIR, proponents say. But getting there will be a long road. In: *Chemical & Engineering News* (97). Online verfügbar unter <https://cen.acs.org/policy/publishing/Chemistry-data-should-FAIR-proponents/97/i35>, zuletzt geprüft am 10.09.2019.

Initiative NFDI4Chem (2019): Binding Letter of Intent der NFDI4Chem. Online verfügbar unter <https://www.dfg.de/foerderung/programme/nfdi/absichtserklaerungen/index.html>, zuletzt geprüft am 08.09.2019.

IUPAC (Hg.) (o. J.): JCAMP-DX. Online verfügbar unter <http://www.jcamp-dx.org/>, zuletzt geprüft am 11.09.2019.

IUPAC (Hg.) (2018): DIGChem – a Vision for Chemical Data Standards. Online verfügbar unter <https://iupac.org/digchem-a-vision-for-chemical-data-standards/>, zuletzt aktualisiert am 13.09.2019.

IUPAC (Hg.) (2019): Committee on Publications and Cheminformatics Data Standards. Online verfügbar unter https://iupac.org/who-we-are/committees/committee-details/?body_code=024, zuletzt aktualisiert am 13.09.2019.

Karlsruhe Institute of Technology, Institute for Organic Chemistry; German Research Foundation (Hg.) (o.J.): Chemotion. Online verfügbar unter <https://www.chemotion.net/>, zuletzt geprüft am 08.09.2019.

Koepler, Oliver; Jung, Nicole; Kraft, Angelina; Neumann, Janna (2018): Thesenpapier Nationale Forschungsdateninfrastruktur Für Die Chemie (Nfdi4Chem), zuletzt geprüft am 11.09.2019.

Lorenz Gräf (2010): Online-Befragung. Eine praktische Einführung für Anfänger. Berlin: LIT Verlag.

Ludwig, Jens; Enke, Harry (Hg.) (2013): Leitfaden zum Forschungsdaten-Management. Handreichungen aus dem WissGrid-Projekt. Göttingen, Glückstadt: Niedersächsische Staats- und Universitätsbibliothek; Hülsbusch, zuletzt geprüft am 13.09.2019.

McEwen, Leah; Chalk, Stuart; Bruno, Ian; Martinsen, David; Kidd, Richard (Hg.) (2019): Chemistry Research Data Interest Group. Online verfügbar unter <https://www.rd-alliance.org/groups/chemistry-research-data-interest-group.html>, zuletzt aktualisiert am 13.09.2019.

Menalib (Hg.) (o.J.): Umfrage: Forschungsdaten in den Orientwissenschaften. Online verfügbar unter <https://www.menalib.de/community/umfragen/umfrage-forschungsdaten-in-den-orientwissenschaften/>, zuletzt geprüft am 13.09.2019.

Mohler, Peter Ph.; Porst, Rolf (1996): Pretest und Weiterentwicklung von Fragebogen - Einführung in das Thema. In: Wiesbaden Statistische Bundesamt (Hg.): Pretest und Weiterentwicklung von Fragebogen, Bd. 9. Stuttgart: Metzler-Poeschel (Spektrum Bundesstatistik, 9), S. 7–15.

Müller, Hendrik; Sedley, Aaron; Ferrall-Nunge, Elizabeth (2014): Survey Research in HCI. In: Judith S. Olson und Wendy A. Kellogg (Hg.): Ways of Knowing in HCI. New York, NY: Springer New York, S. 229–267.

National Metrology Institute of Japan; National Institute of Advanced Industrial Science and Technology (Hg.) (o.J.): Spectral Database for Organic Compounds. Online verfügbar unter https://sdfs.db.aist.go.jp/sdfs/cgi-bin/cre_index.cgi, zuletzt geprüft am 11.09.2019.

NFDI4Chem Online Survey (2019). Online verfügbar unter <https://www.nfdi4chem.de/index.php/2019/09/05/nfdi4chem-online-survey/>, zuletzt geprüft am 13.09.2019.

Philipps-Universität Marburg (Hg.) (o.J.): HeFDi - Hessische Forschungsdateninfrastrukturen. Online verfügbar unter <https://www.uni-marburg.de/de/forschung/kontakt/forschungsdatenmanagement/projekte/hefdi-hessische-forschungsdateninfrastrukturen>, zuletzt geprüft am 15.09.2019.

Pupier, Marion; Nuzillard, Jean-Marc; Wist, Julien; Schlörer, Nils E.; Kuhn, Stefan; Erdelyi, Mate et al. (2018): NMReDATA, a standard to report the NMR assignment and parameters of organic compounds. In: *Magnetic resonance in chemistry : MRC* 56 (8), S. 703–715. DOI: 10.1002/mrc.4737.

Rat für Informationsinfrastrukturen (Hg.) (2016): Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland. Göttingen. Online verfügbar unter <http://www.rfii.de/?p=1998>, zuletzt geprüft am 13.09.2019.

Rat für Informationsinfrastrukturen (Hg.) (2018): Zusammenarbeit als Chance. Zweiter Diskussionsimpuls zur Ausgestaltung einer Nationalen Forschungsdateninfrastruktur (NFDI) für

die Wissenschaft in Deutschland. Göttingen. Online verfügbar unter <http://www.rfii.de/?p=2529>, zuletzt geprüft am 13.09.2019.

re3data.org Project Consortium (Hg.) (o.J.): re3data.org. Online verfügbar unter <https://www.re3data.org/>, zuletzt geprüft am 11.09.2019.

Rühle, Stefanie (2012): Kleines Handbuch Metadaten. Online verfügbar unter http://www.kimforum.org/Subsites/kim/SharedDocs/Downloads/DE/Handbuch/metadaten.pdf?__blob=publicationFile, zuletzt geprüft am 07.09.2019.

Scheuch, Erwin K. (1996): Die Notwendigkeit von Pretests zur Vorbereitung statistischer Erhebungen. In: Wiesbaden Statistische Bundesamt (Hg.): Pretest und Weiterentwicklung von Fragebogen, Bd. 9. Stuttgart: Metzler-Poeschel (Spektrum Bundesstatistik, 9), S. 16–27.

Schweiger, Wolfgang (o.J.): Fragen und Antworten: von der Forschungsfrage zum Fragebogen. Technische Universität Dresden. Online verfügbar unter <https://docplayer.org/3074067-Fragen-und-antworten-von-der-forschungsfrage-zum-fragebogen.html>, zuletzt geprüft am 27.08.2019.

SoSciSurvey GmbH (o.J.): SoSciSurvey - der onlineFragebogen.

Suche Umfrage und Forschungsdatenmanagement (2019). Online verfügbar unter https://www.google.com/search?client=firefox-b-d&sxsrf=ACYBGNQB9H8n7WdvBx0ILMTAJOhd9k7cNw%3A1568384551199&ei=J6Z7XY3lC4XikgWlgl-4Dw&q=forschungsdatenmanagement+umfrage&oq=forschungsdatenmanagement+umfrage&gs_l=psy-ab.3.0i8i30l2.1444904.1445703.1446003...0.0.0.59.103.2.....0...1.gws-wiz.KgCFBcpK2mo&ved=0ahUKEwjNIK7g_83kAhUFsaQKHSXAD_cQ4dUDCAo&uact=5, zuletzt geprüft am 13.09.2019.

Technische Informationsbibliothek; Fachinformationszentrum Chemie GmbH; Universität Paderborn (Hg.) (2010): Konzeptstudie - Forschungsdaten in der Chemie, zuletzt geprüft am 12.09.2019.

The FAIRsharing team; Oxford e Research Centre, University of Oxford (Hg.) (2009-2019): fairsharing.org. Online verfügbar unter <https://fairsharing.org/>, zuletzt geprüft am 13.09.2019.

Thielsch, Meinald T.; Weltzin, Simone (2012): Online-Befragungen in der Praxis. In: Meinald T. Thielsch und Torsten Brandenburg (Hg.): Praxis der Wirtschaftspsychologie II. Münster: MV Wissenschaft, S. 109–127.

Treloar, Andrew; Harboe-Rae, Cathrine (2008): Data management and the curation continuum: how the Monash experience is informing repository relationships. Online verfügbar unter <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.572.4602&rep=rep1&type=pdf>, zuletzt geprüft am 13.09.2019.

Tristram, Frank (2015): Abschlussbericht bwFDM-Communities. Online verfügbar unter https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja&uact=8&ved=2ahUKEwihpd2Di87kAhWNDewKHaWJAZUQFjABegQIABAB&url=https%3A%2F%2Fwww.forschungsdaten.org%2Findex.php%2FUmfragen_zum_Umgang_mit_Forschungsdaten_an_wisse

nschaftlichen_Institutionen&usg=AOvVaw0g_OZrEr1rINucMsopXQfV, zuletzt geprüft am 13.09.2019.

Universität Duisburg-Essen (Hg.) (o.J.): Landesinitiative NFDI der Digitalen Hochschule NRW. Online verfügbar unter <https://www.fdm.nrw/>, zuletzt aktualisiert am 15.09.2019.

University of Leicester (Hg.) (o.J.): Research data. Online verfügbar unter <https://www2.le.ac.uk/services/research-data/rdm/what-is-rdm/research-data>, zuletzt geprüft am 12.09.2019.

Vereinigung für Afrikawissenschaften in Deutschland (Hg.) (2019): Umfrage zum Thema Forschungsdatenmanagement. Online verfügbar unter <https://vad-ev.de/umfrage-zum-thema-forschungsdaten/>, zuletzt geprüft am 13.09.2019.

Wagner, Pia; Hering, Linda: Online-Befragung, S. 661–673.

Wagner, Pia; Hering, Linda (2014): Online-Befragung. In: Nina Baur und Jörg Blasius (Hg.): Handbuch Methoden der empirischen Sozialforschung. Wiesbaden: Springer VS, S. 661–673.

Waltemath, Dagmar; Adams, Richard; Beard, Daniel A.; Bergmann, Frank T.; Bhalla, Upinder S.; Britten, Randall et al. (2011): Minimum Information About a Simulation Experiment (MIASE). In: *PLoS computational biology* 7 (4), e1001122. DOI: 10.1371/journal.pcbi.1001122.

Wendler, Antje (2018): Statistik. Einführung in Methoden der deskriptiven und induktiven Statistik. In: *Schriften des Verbundstudiengangs Wirtschaftsingenieurwesen*. DOI: 10.2314/GBV1022290940.

Wikipedia, Die Freie Enzyklopädie (Hg.) (2019a): EVA-Prinzip. Online verfügbar unter <https://de.wikipedia.org/w/index.php?title=EVA-Prinzip&oldid=191995368>, zuletzt geprüft am 07.09.2019.

Wikipedia, Die freie Enzyklopädie. (Hg.) (2019b): Elektronisches Laborbuch. Online verfügbar unter https://de.wikipedia.org/w/index.php?title=Elektronisches_Laborbuch&oldid=190154110, zuletzt geprüft am 08.09.2019.

Wilde, Mike (o.J.): Rücklaufquoten bei Online-Umfragen: Definition, Berechnung, Einflussparameter & Best Practices. Online verfügbar unter <https://www.questionpro.de/responsequote-ruecklaufquote-online-umfrage/>, zuletzt geprüft am 19.08.2019.

Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, I. Jsbrand Jan; Appleton, Gabrielle; Axton, Myles; Baak, Arie et al. (2016): The FAIR Guiding Principles for scientific data management and stewardship. In: *Scientific data* 3, S. 160018. DOI: 10.1038/sdata.2016.18.

Winkler-Nees, Stefan (2011): Vorwort. In: Stephan Büttner und Hans-Christoph Hobohm (Hg.): Handbuch Forschungsdatenmanagement. Bad Honnef: Bock + Herchen.

Winter, Nina (2019): FDM-Strukturen an NRW-Hochschulen, zuletzt geprüft am 13.09.2019.

wirtschaftslexikon24.com (o.J.). Online verfügbar unter <http://www.wirtschaftslexikon24.com/e/r%C3%BCcklaufquote/r%C3%BCcklaufquote.htm>, zuletzt geprüft am 26.07.2019.

xStudy SE (Hg.) (1997-2019): studieren.de. Online verfügbar unter https://studieren.de/chemie.hochschulliste.t-0.c-22778.s-1.html?&tx_assearchengine_pi1%5Bproperties%5D=all%3A0%3A19, zuletzt geprüft am 24.07.2019.

Yan, Yichao; Robinson, Sophia G.; Sigman, Matthew S.; Sanford; Melanie S. (2019): Supporting Information for: Mechanism-based design of a high-potential catholyte enables a 3.2 V all-organic nonaqueous redox flow battery. In: *Journal of the American Chemical Society*. DOI: 10.1021/jacs.9b07345.s001.

ZB MED - Informationszentrum Lebenswissenschaften; Adam, Beatrix; Lindstädt, Birte (2019): Elektronische Laborbücher im Kontext von Forschungsdatenmanagement und guter wissenschaftlicher Praxis - ein Wegweiser für die Lebenswissenschaften. Online verfügbar unter <https://www.publisso.de/forschungsdatenmanagement/fd-dokumentieren/el-n-wegweiser/>, zuletzt geprüft am 14.09.2019.

Zerr, Konrad (2003): Online-Marktforschung - Erscheinungsformen und Nutzerpotentiale. In: Axel Theobald, Marcus Dreyer und Thomas Starsetzki (Hg.): *Online-Marktforschung. Theoretische Grundlagen und praktische Erfahrungen*. 2. Aufl. Wiesbaden: Gabler Verlag, S. 7–26.

10.Danksagung

Ich möchte allen meinen Dank aussprechen, die zum Gelingen dieser Arbeit beigetragen haben.

Mein besonderer Dank gilt

Frau Dr. Ania López für die ausgezeichnete Unterstützung und Betreuung dieser Arbeit.

Herrn Prof. Dr. Achim Oßwald für die freundliche Übernahme des Zweitgutachten und die ausgezeichnete Betreuung während der gesamten Studienzeit.

Der leitenden Bibliotheksdirektorin Frau Dr. Ulrike Eich und meinen Kolleginnen und Kollegen, Herrn Dr. Dominik Schmitz, Frau Dr. Ute Trautwein-Bruns, Herrn Stephan von der Ropp, Frau Ji-Hyoung Han und Frau Andrea Jaek, die mir bei dienstlichen und privaten Problemen aller Art während der gesamten Studienzeit beiseite standen.

Meinem Mann Florian für einfach alles.

Meinen Kindern Julia, Elias und Jannis, die mit stets ein Lächeln ins Gesicht gezaubert haben.

Meinen Eltern Renate und Gerhard, die durch ihr Vertrauen und ihrem Glauben an mich, mir stets Kraft gegeben haben. Und fürs Babysitten 😊

Es hat mir sehr viel Spaß gemacht, Danke!

11. Anhang

11.1 Einladungsanschreiben

Sehr geehrte/r Frau/Herr Prof. XXX, sehr geehrte wissenschaftliche und nicht-wissenschaftliche Mitarbeitende,

als Chemikerin unterstütze ich seit 2016 das Forschungsdatenmanagement Team der RWTH Aachen.

Im Rahmen meiner Masterarbeit an der TH Köln zum Thema „Forschungsdaten in der Chemie“ erarbeite ich den Umgang mit Forschungsdaten in der Chemie nach den FAIR-Prinzipien. Dieses trägt dazu bei, chemiespezifische Empfehlungen und ggfs. Angebote zu erarbeiten bzw. anzupassen.

Daher bitte ich Sie, an der zirka 12-minütigen [Umfrage](#) teilzunehmen und mich bei der Erstellung meiner Masterarbeit sowie den Prozess des Forschungsdatenmanagements in der Chemie ohne viel Aufwand zu unterstützen. Die [Umfrage](#) ist vom Zeitraum 29.05. bis 14.06.2019 offen und wird mit dem deutschen Tool Soscisurvey.de durchgeführt. Sämtliche Daten werden auf deutschen Servern gespeichert und unterliegen dem deutschen Datenschutz. Mit der Teilnahme erklären Sie sich mit diesen [Bedingungen](#) einverstanden.

Die Umfrage erreichen Sie unter folgendem Link:
<https://www.socisurvey.de/FAIRChem/>.

Das Passwort für die Umfrage lautet: RWTH_Aachen

Zielgruppe der Umfrage sind sowohl Professorinnen und Professoren sowie wissenschaftliche und nicht-wissenschaftliche Mitarbeitende der Chemie in Nordrhein-Westfalen.

Meine Masterarbeit wird auf dem Repositorium [PubLIS Cologne](#) Open Access veröffentlicht.

Für persönliche Fragen stehe [ich](#) Ihnen jederzeit zur Verfügung.

Ich danke Ihnen im Voraus für Ihre Teilnahme und Unterstützung und verbleibe mit freundlichen Grüßen

Daniela Hausen

11.2 Erinnerungsanschreiben

Sehr geehrte/r Frau/Herr Prof. XXX, sehr geehrte wissenschaftliche und nicht-wissenschaftliche Mitarbeitende,

gerne möchte ich Sie nochmal an die ca. 12-minütige Umfrage im Rahmen meiner Masterarbeit „FAIRness der Forschungsdaten in der Chemie“ erinnern und um Ihre Unterstützung bitten. Die Zielgruppen der Umfrage sind Professorinnen und Professoren, wissenschaftliche und nichtwissenschaftliche Mitarbeitende der nordrhein-westfälischen Universitäten.

Die Umfrage erreichen Sie unter folgendem Link:

<https://www.socisurvey.de/FAIRChem/> .

Das Passwort für die Umfrage lautet: RWTH_Aachen

Sie ist bis heute um 24 Uhr für Sie geöffnet.

Bei Fragen stehe ich Ihnen gerne weiterhin zur Verfügung und bedanke mich im Voraus für Ihre Unterstützung.

Ich verbleibe mit freundlichen Grüßen

Daniela Hausen

11.3 Fragebogen

In diesem Anhang finden Sie Screenshots des Fragebogens, um so einen besseren Eindruck vom Layout zu bekommen. Außerdem wurden einige Fragen zusammengezogen, was sich so besser darstellen lässt.

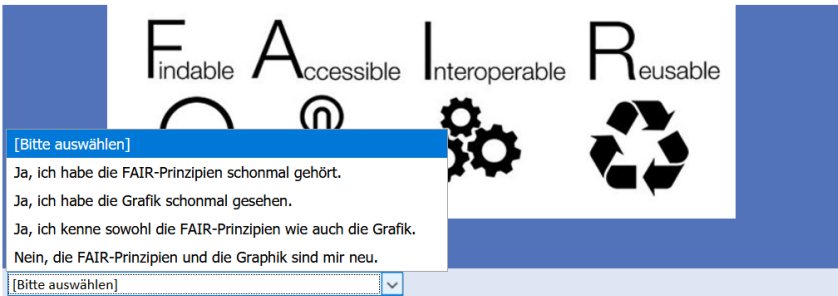
Seite 01

1. Welche Position haben Sie an Ihrem Institut inne?

- Professorin/ Professor
- Postdoc/ Akademische Rätin/ Akademischer Rat/ Oberingenieurin/ Oberingenieur ...
- Doktorandin/ Doktorand
- technische Mitarbeiterin/ technischer Mitarbeiter
- Andere

Seite 02

2. Kennen Sie die FAIR-Prinzipien? Ist Ihnen die folgende Grafik vertraut?



[Bitte auswählen]
Ja, ich habe die FAIR-Prinzipien schonmal gehört.
Ja, ich habe die Grafik schonmal gesehen.
Ja, ich kenne sowohl die FAIR-Prinzipien wie auch die Grafik.
Nein, die FAIR-Prinzipien und die Graphik sind mir neu.
[Bitte auswählen]

Abbildung 32: Fragebogen in SoSciSurvey - Fragen 1 und 2

Seite 03

3. Welche Daten erheben Sie bei Ihrer Forschung?

Mehrfachantworten möglich

- Experimentelle Daten
- Simulationsdaten
- Chromatographische Daten
- Spektroskopische Daten
- Mikroskopische Daten
- Kristallographische Daten
- Ellipsometrische Daten
- Interferometrische Daten
- Topografische Daten
- Elektrochemische Daten
- Andere Daten

Abbildung 33: Fragebogen in SoSciSurvey – Frage 3

4. Welche Versionen Ihrer Daten z.B. Rohdaten, aufgearbeitete, analysierte oder berechnete Daten speichern Sie am Projektende ab?

Mehrfachantworten möglich

Rohdaten

Aufgearbeiteten Daten

Analysierte Daten

Berechnete Daten

Andere Daten

Alle Daten

Ich lösche mit Bedacht die Daten, die ich nicht auswerten kann bzw. von den Experimenten, Messungen, etc., die nicht funktioniert haben.

Abbildung 34: Fragebogen in SoSciSurvey - Frage 4

5. Um Daten innerhalb Ihres Instituts, Ihres Projektes, für eine Publikation, etc. auffindbar zu machen, ist die Beschreibung Ihrer Daten mit "weiterführenden Informationen" sogenannten Metadaten essentiell.

Wenn Sie nun an Ihre Daten denken. Mit welchen Metadaten beschreiben Sie Ihre Daten?

Bitte bewerten Sie diese auf einer Skala von 0 bis 9.

Zu werten dabei ist 0 = unbedeutend und 9 = sehr wichtig.

	ist unbedeutend	ist sehr wichtig	kann ich nicht beurteilen								
	0 1 2 3 4 5 6 7 8 9										
Strukturformel	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Summenformel	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Analysemethode	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Metadaten: <input type="text"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Metadaten: <input type="text"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Metadaten: <input type="text"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Metadaten: <input type="text"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Metadaten: <input type="text"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Metadaten: <input type="text"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Metadaten: <input type="text"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Metadaten: <input type="text"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Metadaten: <input type="text"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Metadaten: <input type="text"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Abbildung 35: Fragebogen in SoSciSurvey - Frage 5

6. Könnte eine Wissenschaftlerin/ ein Wissenschaftler innerhalb oder außerhalb Ihres Instituts mit den bisher von Ihnen mitgelieferten Metadaten, Ihre Daten verstehen?
Wenn Sie sich nicht sicher sind, fragen Sie Ihre Kollegin/ Ihren Kollegen, ob sie/ er die Daten versteht

Ja, innerhalb meines Institutes, meiner Arbeitsgruppe,

Ja, innerhalb und außerhalb meines Institutes, meiner Arbeitsgruppe,

Ja, allerdings nur den Teil meiner Daten, die zur Veröffentlichung aufbereitet wurden.

Ich denke nicht, dass eine andere Wissenschaftlerin oder ein anderer Wissenschaftler meine Daten ohne weitere Erläuterungen meinerseits verstehen kann.

7. Was für eine Art Laborbuch (Papier oder elektronisches Laborbuch) verwenden Sie? Welche Gründe sprechen für die jeweilige Nutzung?

Ich verwende ein .

Ich verwende es, .

- Laborbuch aus Papier
- elektronisches Laborbuch
- Ich benutze kein Laborbuch.

Abbildung 36: Fragebogen in SoSciSurvey - Fragen 6 und 7

8. Neben den Metadaten ist auch die Strukturierung ihrer Daten ein wichtiges Element. Wie strukturieren bzw. benennen Sie ihre Daten?
Mehrfachantworten möglich

Mögliche Elemente sind:

- Datum der Erstellung
- Summenformel nach IUPAC
- CAS-Nummer
- Projektnummer
- Name des Forschers
- Methode der Datenerhebung
- Versionsnummer
- Nutzungsrechte
- Andere
- Andere
- Andere

Abbildung 37: Fragebogen in SoSciSurvey - Frage 8

9. Strukturieren und/ oder beschreiben Sie Ihre Daten mit Metadaten nach bestimmten Vorgaben, Absprachen, usw.?
Mehrfachantworten möglich

- Ja, in meiner Teildisziplin gibt es einheitliche Vorgaben zur Strukturierung.
- Ja, an meinem Institut gibt es einheitliche Vorgaben zur Strukturierung.
- Ja, in meiner Arbeitsgruppe gibt es einheitliche Vorgaben zur Strukturierung.
- Ja, in meiner Teildisziplin gibt es einheitliche Vorgaben zur Erfassung der Metadaten.
- Ja, an meinem Institut gibt es einheitliche Vorgaben zur Erfassung der Metadaten.
- Ja, in meiner Arbeitsgruppe gibt es einheitliche Vorgaben zur Erfassung der Metadaten.
- Ja, ich habe mich mit meinen Kolleginnen und Kollegen abgesprochen.
- Nein, es gibt keine Vorgaben zur Strukturierung.
- Nein, es gibt keine Vorgaben zur Erfassung der Metadaten.
- Nein, die gemachten Vorgaben machen für mich keinen Sinn. Ich habe mein eigenes System der Strukturierung und erfasse die Metadaten, die ich für meine Daten relevant halte.

Abbildung 38: Fragebogen in SoSciSurvey - Frage 9

10. Wie teilen Sie Daten innerhalb Ihres Instituts?
Mehrfachantworten möglich

- elektronisches Laborbuch
- Institutsserver
- E-Mail
- USB-Stick
- SharePoint
- Sciebo
- DropBox
- BSCW
- GitLab
- Ich teile meine Daten nicht.
- Andere

Abbildung 39: Fragebogen in SoSciSurvey - Frage10

11. Wie teilen Sie Daten außerhalb Ihres Instituts?
Mehrfachantworten möglich

- elektronisches Laborbuch
- E-Mail
- USB-Stick
- SharePoint
- Sciebo
- DropBox
- BSCW
- GitLab
- Veröffentlichung
- Ich teile meine Daten nicht.
- Andere

12. Haben Sie bereits Daten in Form von Rohdaten, bearbeiteten Daten, ... veröffentlicht?
Mehrfachantworten möglich

- Ja, im Rahmen einer Textpublikation als supplement material im pdf-Format.
- Ja, im Rahmen einer Textpublikation auf einem Repositorium.
- Ja, unabhängig von einer Textpublikation als reine Datenpublikation auf einem Repositorium.
- Nein, ich habe bisher noch keine Daten veröffentlicht.
- Andere

Abbildung 40: Fragebogen in SoSciSurvey - Fragen 11 und 12

13. Haben Sie selbst schonmal Daten aus anderen Quellen nachgenutzt?
Mehrfachantworten möglich

- Ja, Daten von einer Kollegin/ einem Kollegen aus der eigenen Arbeitsgruppe/ dem eigenen Institut.
- Ja, Daten von einer Kollegin/ einem Kollegen aus einer anderen Arbeitsgruppe/ einem anderen Institut.
- Ja, Daten aus den supplement materials.
- Ja, Daten, die auf einem Repositorium zur Verfügung standen.
- Nein, Daten waren nicht ausreichend beschrieben.
- Nein, ich vertraue fremden Daten nicht.
- Nein, bisher gab es noch nicht die Notwendigkeit fremde Daten nachzunutzen.
- Nein, diese Idee hatte ich noch nicht.

14. Vielen Dank für Ihre Teilnahme!
Ich möchte mich ganz herzlich für Ihre Mithilfe an meiner Masterarbeit bedanken.
Bei allgemeinen und persönlichen Rückfragen oder Fragen zum Thema Forschungsdatenmanagement stehe ich Ihnen gerne zur Verfügung.
Daniela Hausen, hausen@ub.rwth-aachen.de, +49 241 80 90314

Für aktuelle Fragen, Bemerkungen und Anregungen finden Sie hier Platz:

Abbildung 41: Fragebogen in SoSciSurvey - Frage 13 und Anmerkungen

Vielen Dank für Ihre Teilnahme!

Ihre Antworten wurden gespeichert, Sie können das Browser-Fenster nun schließen.

Dipl.Chem. Daniela Hausen, RWTH Aachen University – 2019

Abbildung 42: Fragebogen in SoSciSurvey - Danksagung

12 Eidesstaatliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe.

Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Dies gilt auch für Quellen aus eigenen Arbeiten.

Ich versichere, dass ich diese Arbeit oder nicht zitierte Teile daraus vorher nicht in einem anderen Prüfungsverfahren eingereicht habe.

Mir ist bekannt, dass meine Arbeit zum Zwecke eines Plagiatsabgleichs mittels einer Plagiatserkennungssoftware auf ungekennzeichnete Übernahme von fremdem geistigem Eigentum überprüft werden kann.

Jülich, den 16.09.2019

