

HUMBOLDT-UNIVERSITÄT ZU BERLIN
INSTITUT FÜR BIBLIOTHEKS- UND INFORMATIONSWISSENSCHAFT



BERLINER HANDREICHUNGEN
ZUR BIBLIOTHEKS- UND
INFORMATIONSWISSENSCHAFT

HEFT 455

METADATEN FÜR FORSCHUNGSDATEN

BEDÜRFNISSE UND ANFORDERUNGEN IN DEN
NATURWISSENSCHAFTEN

VON
TEREZA KALOVÁ

METADATEN FÜR FORSCHUNGSDATEN
BEDÜRFNISSE UND ANFORDERUNGEN IN DEN
NATURWISSENSCHAFTEN

VON
TEREZA KALOVÁ

Berliner Handreichungen zur
Bibliotheks- und Informationswissenschaft

Begründet von Peter Zahn
Herausgegeben von
Vivien Petras
Humboldt-Universität zu Berlin

Heft 455

Kalová, Tereza

Metadaten für Forschungsdaten : Bedürfnisse und Anforderungen in den Naturwissenschaften / von Tereza Kalová. - Berlin : Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin, 2020. – 74 S. : graph. Darst. - (Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft ; 455)

ISSN 14 38-76 62

Abstract:

Bei der Entwicklung von bedarfsgerechten Services im Bereich eScience sind Untersuchungen der Bedürfnisse und Anforderungen der Zielgruppen von besonderer Relevanz. Angelehnt an die Grounded Theory wurden im Rahmen des österreichweiten Projektes e-Infrastructures Austria Plus von April bis Mai 2019 acht Leitfadeninterviews mit Naturwissenschaftlern von der Universität Wien und der Medizinischen Universität Wien durchgeführt. Die Ergebnisse zeigen, dass ein Verständnis des Begriffes „Metadaten“ unter Forschenden noch nicht flächendeckend vorhanden ist. Es konnten aber mehrere Beispiele der Beschreibung von Daten in der Praxis identifiziert, sowie sechs Bereiche an Wünschen der Forschenden herausgearbeitet werden. Diese beschäftigen sich u. a. mit der Notwendigkeit, Services auf Deutsch und Englisch zu konzipieren, mit den Anforderungen an technische Infrastruktur oder dem Wunsch nach dem Einsatz von Data Stewards, die Fachkenntnisse aus einer relevanten Disziplin mitbringen.

Diese Veröffentlichung geht zurück auf eine Masterarbeit im weiterbildenden Masterstudiengang im Fernstudium Bibliotheks- und Informationswissenschaft (Library and Information Science, M. A. (LIS)) an der Humboldt- Universität zu Berlin.

Eine Online-Version ist auf dem edoc Publikationsserver der Humboldt-Universität zu Berlin verfügbar.



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

Inhaltsverzeichnis

1	Einleitung	8
2	Forschungsdaten, Metadaten und die Bibliothek	10
2.1	Forschungsdaten	11
2.2	Metadaten.....	12
2.3	Forschungsdatenmanagement und Bibliothek	13
2.4	e-Infrastructures Austria und e-Infrastructures Austria Plus	15
2.5	Metadaten und Forschende	16
3	Methodik	19
3.1	Grounded Theory	19
3.2	Das Leitfaden-Interview	21
3.3	Sampling und Feldzugang	23
3.4	Pre-Test	25
3.5	Durchführung der Interviews	25
3.6	Transkribierung.....	26
3.7	Auswertung.....	26
3.7.1	Memos	26
3.7.2	Kodieren (Coding)	27
3.8	Ethik.....	28
4	Die Ergebnisse der qualitativen Interviews	31

4.1	Porträts der Wissenschaftler	31
4.2	Forschungsdaten.....	37
4.3	Metadaten	39
4.3.1	Begriff „Metadaten“	39
4.3.2	Beschreibung von Forschungsdaten in der Praxis	40
4.4	Bedürfnisse und Anforderungen.....	46
4.4.1	Unterschiede zwischen Disziplinen	46
4.4.2	Aufwand.....	46
4.4.3	Sprache	47
4.4.4	Informationsvermittlung.....	48
4.4.5	Personal	50
4.4.6	Technische Infrastruktur	51
5	Diskussion	53
6	Fazit.....	58
	Abbildungsverzeichnis	60
	Literaturverzeichnis	61
	Anhang I – Leitfaden	69
	Anhang II – Codebuch	70
	Anhang III – Transkripte der Interviews	74

1 Einleitung

„If you just collect the primary data and they are not annotated properly, this information is lost, basically.“ (Interview Wissenschaftler C 2019, A. 17)

Metadaten stellen einen der gängigsten Begriffe der Bibliothekswissenschaft dar. Sie kommen sowohl bei Publikationen als zunehmend auch bei Forschungsdaten zum Einsatz. Die Erstellung von Metadaten und Dokumentation von Forschungsdaten bietet zahlreiche Vorteile und eröffnet Wege zu interdisziplinärer Forschungstätigkeit, sowie der Nachnutzung von Forschungsdaten. Das Bewusstsein für das Thema ist allerdings unter Forschenden nicht flächendeckend vorhanden.

Naturwissenschaftler_innen als Zielgruppe stellen sich in diesem Zusammenhang vonseiten der wissenschaftlichen Bibliotheken, als besonders schwer zu erreichen dar, da Bibliotheken traditionell eher mit Geistes- und Sozialwissenschaften in Verbindung gebracht werden. Die vorliegende Arbeit bemüht sich, diese Distanz zu überbrücken, indem sie die Wünsche der Forschenden untersucht.

Die zentrale Fragestellung der Arbeit, nämlich die nach den konkreten Bedürfnissen und Anforderungen der Naturwissenschaftler_innen im Bereich Metadaten für Forschungsdaten, wurde in drei Untersuchungsfelder ausdifferenziert. So wird neben den Wünschen das Verständnis und Bekanntheit des Begriffs „Metadaten“ thematisiert, sowie die Ansätze der Beschreibung von Forschungsdaten in der Praxis abgefragt:

1. Ist der Begriff „Metadaten“ in den Naturwissenschaften bekannt?
2. Wie werden Forschungsdaten in der Praxis beschrieben?
3. Welche Bedürfnisse und Anforderungen gibt es in Bezug auf Metadaten für Forschungsdaten in den Naturwissenschaften? Welche Services werden benötigt?

Um diese Fragen zu beantworten, wurden qualitative Interviews mit Forschenden durchgeführt.

Die Arbeit wird in sechs Kapitel aufgeteilt. Im Kapitel zwei werden die relevanten Begriffe „Forschungsdaten“ und „Metadaten“ definiert, sowie die Rolle von Bibliotheken im Forschungsdatenmanagement geschildert. Weiter wird auf die Ergebnisse bisheriger Studien zu Bedürfnissen und Anforderungen von Forschenden im Bereich Metadaten eingegangen.

Kapitel drei beschäftigt sich mit der angewandten Methodik, nämlich den wesentlichen Aspekten der *Grounded Theory*, Leitfaden-Interviews, sowie der Durchführung von Interviews. Außerdem wird der Prozess der Auswertung mithilfe des *Codings* und analytischer Memos erklärt.

Im nächsten Kapitel werden die Ergebnisse der qualitativen Interviews thematisch geordnet vorgestellt. Zunächst werden die einzelnen Teilnehmenden skizziert. Weiter wird das Verständnis des Begriffs Metadaten ausgearbeitet. Es werden praktische Ansätze in der Beschreibung von Forschungsdaten dargestellt, sowie die Bedürfnisse und Anforderungen der Befragten ausformuliert.

In der Diskussion werden die Ergebnisse nochmals zur Beantwortung der Forschungsfragen mit weiteren Kommentaren herangeführt. So werden potenzielle weitere Untersuchungsfelder identifiziert und Handlungsbedarf für Bibliotheken angeschnitten. Das Fazit fasst anschließend die gesamte Arbeit zusammen.

Die Arbeit wurde im Rahmen des österreichischen Projektes e-Infrastructures Austria Plus durchgeführt. Teilergebnisse der vorliegenden Arbeit wurden im September 2019 in Form eines Projektberichtes im Repositorium der Universität Innsbruck veröffentlicht¹.

¹ Der Projektbericht ist unter dem folgenden Link verfügbar: <https://doi.org/10.25651/1.2019.0020>

2 Forschungsdaten, Metadaten und die Bibliothek

Seit mehreren Jahren wird von einer Flut an Daten oder *Big Data* gesprochen. Durch die immer wachsende Digitalisierung steigen auch in der Forschung die Massen an generierten Daten besonders in naturwissenschaftlichen Disziplinen exponentiell an (vgl. Hey und Trefethen 2003, S. 2; Büttner, Hobohm, und Müller 2011, S. 13). Darüber hinaus ist die wachsende Internationalisierung und Interdisziplinarität der Wissenschaft und das damit verbundene internationale kollaborative Arbeiten bei Fragen des richtigen Umgangs mit Forschungsdaten von erheblicher Relevanz. Es wird diesbezüglich von dem s.g. *fourth paradigm* gesprochen (vgl. Hey, Tansley und Tolle 2009). Dies bedeutet, dass Daten aus verschiedenen Quellen und Disziplinen in Verbindung gesetzt werden und diese mit maschineller Unterstützung untersucht werden, um neue Hypothesen zu generieren (vgl. Hey, Tansley und Tolle 2009). Solche datengetriebene Forschung wird in vielen Disziplinen bereits praktiziert und zeigt auch die zukünftige Entwicklung der Forschung vor allem in den Naturwissenschaften auf (vgl. Rapport und Braithwaite 2018, S. 3-6; Zook et al. 2017, S. 1).

An Forschungseinrichtungen inklusive wissenschaftlicher Bibliotheken etablieren sich neue Rollen wie *Data Librarian* oder *Data Manager*, um passende Infrastrukturen für das Management von Forschungsdaten zu entwickeln (Büttner, Rümpel, und Hobohm 2011, S. 206). Diese sollten den Wissenschaftler_innen in erster Linie den Umgang mit digitalen Daten erleichtern, diese nachhaltig für die Zukunft aufzubewahren und für die weitere Nutzung aufzubereiten. Weiter spielen die Ziele von *Open Science*, bzw. *Open Data* eine wichtige Rolle, um die Entwicklung im Bereich Forschungsdatenmanagement voranzutreiben. Da es sich hierbei um ein sehr umfangreiches Thema handelt, das in der wissenschaftlichen Literatur in den letzten Jahren umfassend untersucht wurde, werden in diesem Kapitel daher vor allem die Grundbegriffe „Forschungsdaten“ und „Metadaten“ definiert, die Rolle von Bibliotheken in Forschungsdatenmanagement-Services illustriert, sowie der Stand der Forschung zu Bedürfnissen und Anforderungen an Metadaten für Forschungsdaten unter Naturwissenschaftler_innen dargestellt.

2.1 Forschungsdaten

Forschungsdaten einheitlich zu definieren ist keine einfache Aufgabe. Allgemein lassen sich digitale Forschungsdaten als „alle digital vorliegende[] Daten, die während des Forschungsprozesses entstehen oder ihr Ergebnis sind“ bezeichnen (Kindling und Schirnbacher 2013, S. 130). Konkreter definiert die deutsche Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen Forschungsdaten als „Daten, die im Zuge wissenschaftlicher Vorhaben z.B. durch Digitalisierung, Quellenforschungen, Experimente, Messungen, Erhebungen oder Befragungen entstehen“ (Deutsches GeoForschungsZentrum o.D). Wie viele weitere Quellen orientiert sich diese Definition an der Art der Erhebung oder Erzeugung von Forschungsdaten und stellt dadurch die große Vielfalt von Forschungsdaten dar. Die Bedeutung des Begriffs wird darüber hinaus durch die Heterogenität verschiedener Disziplinen und wissenschaftlicher Methoden geprägt und je nach Fachgebiet unterschiedlich verwendet (vgl. Deutsche Forschungsgemeinschaft 2015).

Abgesehen von der Vielfalt an Bedeutungen, die der Begriff „Forschungsdaten“ umfasst, einigen sich Expert_innen, dass digitale Forschungsdaten wichtige Ergebnisse wissenschaftlicher Tätigkeiten sind, „in deren Erstellung viel Geld, Zeit und Expertise investiert wird. Gleichzeitig sind durch die modernen wissenschaftlichen Arbeitsinstrumente das Volumen und die Komplexität der Forschungsdaten gestiegen und der sinnvolle Umgang mit Forschungsdaten ist deutlich anspruchsvoller geworden“ (Ludwig und Enke 2013, S. 13). Forschungsdaten sind wertvolle Ereignisse, die weit über den Rahmen des ursprünglichen Projektes, für das sie erzeugt wurden, für weitere Forschung von großer Bedeutung sind. „Der gesamte Prozess, der die Allokation, die Generierung, die Bearbeitung und Anreicherung, die Archivierung und Veröffentlichung von digitalen Forschungsdaten [...] unterstützt, wird [...] unter dem Begriff „Forschungsdatenmanagement“ zusammengefasst“ (Kindling und Schirnbacher 2013, S. 130).

Die Gründe, nachhaltiges Forschungsdatenmanagement zu betreiben, sind mannigfaltig. Den Anforderungen der Fördergeber wie der Europäischen Kommission oder in Österreich dem Wissenschaftsfonds FWF zu entsprechen, die

zunehmenden Publikationsbedingungen wissenschaftlicher Verlage sowie die Nachvollziehbarkeit und Reproduzierbarkeit der Ergebnisse ist nur eine kleine Auswahl. Wilkinson et al. zufolge, *„[g]ood data management is not a goal in itself, but rather is the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse“* (2016, S. 1). Forschungsdatenmanagement ermöglicht die Generierung neuer Hypothesen und führt zu wissenschaftlicher Erkenntnis im Sinne des *fourth paradigm*, womit es einen wichtigen Beitrag für die Wissenschaft von Heute und Morgen leistet.

2.2 Metadaten

Ähnlich zu den verschiedenen Definitionen von Forschungsdaten, ist auch der Begriff „Metadaten“ nicht leicht einzugrenzen. Die gängige Definition als „Daten über Daten“ ist zwar etymologisch korrekt, führt aber ohne weitere Erläuterung nicht zum konkreten Verständnis. In bibliothekswissenschaftlichem Kontext wird der Begriff seit Jahren für Informationen über Bücher und Zeitschriften mit Angaben wie Autor oder Seitenzahl verwendet. Metadaten kommen weiter z.B. bei digitalen Objekten in Repositorien zum Einsatz. Jensen, Katsanidou und Zenk-Möltgen definieren „Metadaten [als] Daten oder Informationen, die in strukturierter Form analoge oder digitale Forschungsdaten (Objekte) dokumentieren“ (2011, S. 83). Die Nachnutzung von Forschungsdaten fordert strukturierte Informationen in Form von Metadaten, die „das Finden, Durchsuchen und Nutzen von Forschungsdaten“ ermöglichen (forschungsdaten.org o.D., Abs. 1). Metadaten können in verschiedene Typen unterteilt werden. Es handelt sich beispielsweise um deskriptive und administrative Metadaten (wie Autor) oder technische Metadaten (wie Filetyp) (vgl. Riley und NISO 2017, S. 7).

Bei Forschungsdaten spielen Metadaten eine entscheidende Rolle, FAIR Data Prinzipien zu erfüllen. Dies bedeutet: *„all research objects should be Findable, Accessible, Interoperable and Reusable (FAIR) both for machines and for people“* (Wilkinson et al. 2016, S. 3). Die FAIR Prinzipien werden von vielen Instanzen, z.B. der Europäischen Kommission, als Leitlinien für nachhaltiges Forschungsdatenmanagement eingesetzt. Die FAIR Prinzipien beziehen sich auf die richtige Aufbereitung von Forschungsdaten. Metadaten sind bei jedem der vier Attribute von FAIR Data von größter Relevanz. Beispielsweise sollen

Forschungsdaten mit „*rich metadata*“ versehen werden, um sie Auffindbar (*Findable*) zu machen (GO Fair 2019, A. 2). Strukturierte Metadaten ermöglichen, dass sowohl Menschen als auch Maschinen, die diese Daten nicht selber erzeugt haben, Daten verstehen können und diese demnach auch interdisziplinär verwenden können (vgl. Büttner, Hobohm, und Müller 2011, S. 7). Weiter ist auch die „Datendokumentation essentiell, die den Kontext der Forschungsdaten beschreib[t]“ (Kindling und Schirmbacher 2013, S. 130). Dokumentation in diesem Sinne umfasst alle zusätzlichen relevanten Informationen über den Kontext der Entstehung der Daten, die für die (Nach-)nutzung eine Rolle spielen. Blumesberger argumentiert in diesem Zusammenhang: „Ohne Metadaten ist [ein] Objekt mehr oder weniger wertlos“ (2015, S. 517). Um den Anforderungen an Metadaten aufgrund der interdisziplinären Unterschiede Rechnung zu tragen, wurden viele Metadatenstandards wie *Dublin Core*² entwickelt, bzw. haben sich bereits etabliert (vgl. Research Data Alliance o.D.).

2.3 Forschungsdatenmanagement und Bibliothek

Um Forschende bei der komplexen Aufgabe des richtigen Umgangs mit Forschungsdaten zu unterstützen, entwickeln verschiedene Instanzen an Forschungseinrichtungen bereits seit Jahren Services im Bereich Forschungsdatenmanagement. Wie aus der internationalen Untersuchung von Matusiak und Sposito bezüglich FDM-Services in Europa, Nordamerika und Australien deutlich, sind diese überwiegend an wissenschaftlichen Bibliotheken angesiedelt (vgl. 2017, S. 755). Bibliotheken setzten sich stark mit dem Thema „als konsequente Weiterentwicklung ihrer traditionellen Aufgaben Sammeln, Vermitteln, Bewahren“ auseinander (Bibliotheksdienst 2016, S. 614). Forschungsdatenmanagement bietet Bibliotheken die Möglichkeit eines neuen spannenden Tätigkeitsfeldes. Dies wird allerdings von manchen ebenso als Herausforderung betrachtet, indem Bibliotheken deren eher passive Rolle des Aufbewahrens von Forschungsergebnissen umdenken müssen, um aktive Partner der Wissenschaftler_innen während des gesamten Forschungsprozesses zu werden (vgl. Morgan, Duffield und Walkley Hall 2017, S. 304; Martin 2013, S. 20).

² Siehe <http://www.dublincore.org/specifications/dublin-core/>

Tenopir et al. haben 2017 in einem Survey mit Bibliotheksdirektor_innen die Services im Bereich FDM an wissenschaftlichen Bibliotheken in Europa untersucht. Sie haben gezeigt, dass die überwiegende Mehrheit der Bibliotheken Services im Bereich FDM bereits anbietet oder sind diese in Planung. Die Studie unterscheidet zwischen zwei Gruppen an angebotenen Services, nämlich Beratungsangeboten und technischen Services. Die Untersuchung legt dar, dass die überwiegende Mehrheit der FDM-Services in den Bereich Informations- und Beratungsangebote wie das Erstellen von Info-Websites oder Beratung von Forschenden fällt (Tenopir et al. 2017, S. 2). Die Führungskräfte von Bibliotheken sind sich zudem einig, dass Forschungsdatenmanagement in den Tätigkeitsbereich von Bibliotheken gehört, um die Relevanz der Einrichtungen langfristig zu sichern (Tenopir et al. 2017, S. 5). Weiter kann der Trend zur Annahme des Themas FDM durch Bibliotheken an dem wachsenden Anteil der relevanten Kurse im bibliothekswissenschaftlichen Studium (vgl. Letsch 2017, S. 124) beobachtet werden. Zudem entwickelten sich in den letzten Jahren im deutschsprachigen Raum viele neue Studiengänge und Zertifikatskurse zu *Data Stewardship* und *Data Librarianship*³.

Obwohl Forschungsdatenmanagement als eine logische Weiterentwicklung von traditionellen bibliothekarischen Tätigkeiten betrachtet wird und Bibliothekspersonal den Forschenden beispielsweise im Bereich Metadaten deren Expertise anbieten können, gibt es vonseiten der Wissenschaftler_innen Vorbehalte. „*There can be barriers for researchers to accept librarians in this role,*“ so Surkis und Read von der *NYU Health Sciences Library* (2015, S. 156). Sie schlagen in deren Artikel zum FDM an Bibliotheken vor, dass Bibliothekar_innen im Gespräch mit Forschenden unbedingt die Sprache der Forschung statt des Bibliothekswesens verwenden sollen (vgl. Surkis und Read 2015, S. 156). Vielleicht auch deshalb sind Bibliotheken nicht die einzigen Stakeholder an Services im Bereich Forschungsdatenmanagement. Es gibt bereits mehrere kommerzielle Lösungen für Forschungsdatenmanagement wie *Figshare*, die sehr

³ Siehe zum Beispiel

HU Berlin und FH Potsdam: <https://www.ibi.hu-berlin.de/de/studium/studiengaenge/ddm-master>

Universität Wien: <https://www.postgraduatecenter.at/weiterbildungsprogramme/kommunikation-medien/data-librarian/>

TH Köln: https://www.th-koeln.de/hochschule/neuer-zertifikatskurs-data-librarian-des-zbiw-der-th-koeln-fuer-beschaefigte-wissenschaftlicher-bibliotheken_66083.php

nutzerorientierte Services anbieten und eine Konkurrenz zu Bibliotheken darstellen (vgl. Fingerle 2015). Weiter können sich auch andere Servicestellen wie Datenzentren, die besondere Fachkenntnisse aus den jeweiligen Disziplinen mitbringen, als Akteure in Forschungsdatenmanagement einbringen (vgl. Martin 2013, S. 17). Bibliotheken können hingegen den Vorteil des Vertrauens seitens der Forschenden, sowie der traditionellen und geografischen Nähe zu der *Faculty* genießen (vgl. Martin 2013, S. 12, 19), um bedarfsgerechte Angebote im Bereich Forschungsdatenmanagement zu entwickeln.

2.4 e-Infrastructures Austria und e-Infrastructures Austria Plus

Um den Anforderungen moderner Wissenschaft Rechnung zu tragen und den strategischen Aufbau der entsprechenden bedarfsorientierten Infrastrukturen für eScience an österreichischen Forschungseinrichtungen einzuleiten, hat das österreichische Bundesministerium für Wissenschaft, Forschung und Wirtschaft 2014 das Projekt „e-Infrastructures Austria“, das von der Universität Wien geleitet wurde, gestartet (Universität Wien o.D.b, Abs. 1).

Als eins der Hauptziele des Projektes wurde 2015 eine österreichweite Befragung von Forschenden zum Umgang mit Forschungsdaten unter dem Titel *Forschende und ihre Daten* durchgeführt. Im Rahmen der Umfrage wurden mehr als 3000 Wissenschaftler_innen von allen 21 öffentlich-finanzierten Universitäten und drei außeruniversitären Forschungseinrichtungen befragt. Dies entspricht einer Rücklaufquote von neun Prozent (Bauer et al. 2015, S. 18). Das Ziel der Umfrage war vor allem den Status Quo in Österreich abzubilden, um anhand dessen neue Angebote an den jeweiligen Einrichtungen im Bereich eScience und Forschungsdatenmanagement aufzubauen. Die Studie hat u.a. gezeigt, dass 89 % der Forschenden sich unterstützende Services im Bereich FDM wünschen (Bauer et al. 2015, S. 67). Dies betont den klaren Bedarf an Entwicklung solcher Services in Österreich.

Das Projekt „e-Infrastructures Austria Plus“ ist die Fortsetzung des ersten Projektes. Die Laufzeit war von Herbst 2017 bis Herbst 2019 und die Koordination war an der Universität Innsbruck angesiedelt. An dieser Runde des Projektes nahmen neun öffentlich finanzierte Universitäten und sechs assoziierte Partner

teil. Das Ziel des Projektes war es weiter „Infrastruktur für eScience in Österreich auf[zu]bauen“ (Universität Innsbruck o.D.). Das Projekt wurde in acht Arbeitspakete, die sich mit unterschiedlichen Aspekten von eScience beschäftigen, aufgeteilt. Im Rahmen dieser erarbeiteten Mitarbeiter und Mitarbeiterinnen der Partnereinrichtungen unter den Begriffen „Research Lifecycle“, „FDM Policy“, „Datenmanagementpläne“, „Institutionelle Repositorien“, „Metadaten“, „Persistente Identifikatoren“, „Lab Notebook“ und „Go Fair“ kooperativ miteinander Lösungen für nachhaltiges Forschungsdatenmanagement.

Das Ziel des Vorgängerprojekts war u.a. den Umgang von Forschenden mit Forschungsdaten an österreichischen Einrichtungen wissenschaftlich darzustellen. Aufbauend darauf wurden im Rahmen des Nachfolgeprojektes qualitative Studien an den einzelnen Universitäten durchgeführt, um mehr über die spezifischen Bedürfnisse und Anforderungen der Wissenschaftler_innen in Bezug auf Forschungsdatenmanagement herauszufinden. Bedauerlicherweise sind diese derzeit noch nicht einsehbar. Die vorliegende Arbeit wurde im Rahmen des Arbeitspakets „Metadaten“ des Projektes durchgeführt.

2.5 Metadaten und Forschende

Trotz der zahlreichen Gründe für Forschungsdatenmanagement, inklusive der Beschreibung von Forschungsdaten in Form von Metadaten und Dokumentation, die wissenschaftliche Literatur belegt, müssen bei der Entwicklung von FDM-Services an erster Stelle die Wünsche der Forschenden an der eigenen Einrichtung anerkannt werden. Im Rahmen der Entwicklung neuer Services wird häufig eine Bedarfserhebung mittels qualitativer Interviews oder Online-Surveys realisiert. Sie zielen darauf ab, den Bedarf der *Faculty* möglichst realitätsnah abzubilden, sowie deren Kenntnisse im Bereich FDM abzufragen. Das Thema Metadaten wird häufig, obgleich nicht immer, in Studien thematisiert. Da es sich um einen Aspekt des FDMs handelt, werden Metadaten generell nur als eines der Themen in Form von einer Frage behandelt. Weiter ist wichtig zu bemerken, dass es sich hier in der Regel um pauschale nicht fachspezifische Studien, die den Bedarf an ganzen Einrichtungen abbilden sollen, handelt. In diesem Abschnitt werden demnach von ausgewählten Studien Ergebnisse, die sich konkret auf Metadaten und/oder Dokumentation von Forschungsdaten beziehen, aufgeführt.

Der Bericht vom Projekt e-Infrastructures erwähnt, dass viele Forschende deren Daten beschreiben, dies allerdings sehr uneinheitlich (vgl. Bauer et al. 2015, S. 31). Weiter erwähnt die Studie einen Bedarf an Informationsvermittlung in Form von Beratung und Schulungen (vgl. S. 67). In einer Studie an der *University of Warwick* wurden acht Interviews mit Wissenschaftler_innen, die im Bereich der Physik und Sozialforschung tätig sind, durchgeführt. Einer der Ergebnisse war, dass im Vergleich mit den Sozialforschenden, Physiker_innen Metadaten als eher unwichtig betrachten (vgl. Delasalle 2013, S. 101). Beide Gruppen waren sich aber einig bei deren Vorbehalten gegenüber der Bürokratisierung der Wissenschaft, die sie mit FDM verbinden (vgl. Delasalle 2013, S. 101). Delasalle schlägt diesbezüglich vor, dass das wissenschaftliche Personal in die Entwicklung von Services möglichst einbezogen werden sollte (vgl. 2013, S. 103).

In einer Online-Umfrage an der Universität Trier, an der knapp 200 Forschende teilgenommen haben, (Anteil Humanmedizin/Naturwissenschaften n=51) (vgl. Lemaire et al. 2016, S. 6) hat die Frage nach Datenarchivierung gezeigt, dass der Unterschied zwischen Datensicherung und Datenarchivierung nicht sehr verbreitet ist. Metadaten zu Forschungsdaten werden von Lemaire et al. als der wichtigste Faktor genannt, der die Nachnutzung archivierter Daten ermöglicht (vgl. Lemaire et al. 2016, S. 20). So zeigt die Studie ein mangelndes Bewusstsein für Metadaten. Außerdem belegt die Untersuchung den Bedarf an Informationsvermittlung in diesem Bereich (vgl. Lemaire et al. 2016, S. 20). Weiter schlagen Lemaire et al. vor, dass, um eine Kultur von *Data Sharing* voranzubringen, „Arbeitsprozesse und Werkzeuge so gestaltet sein [müssen], dass bspw. die Beschreibung der Forschungsdaten (Metadaten) keinen übermäßigen Mehraufwand bedeuten“ (2016, S. 28).

Die Umfrage an der Technischen Universität Hamburg (Rücklaufquote 12,8 %, n=96) hat ebenso nachgewiesen, dass sich Forschende Beratung bei technischen Fragen inklusive Metadaten wünschen (vgl. Feldsien-Sudhaus und Rajski 2016, S. 35). In der Befragung der Philipps-Universität Marburg (Beteiligung 18,3 %, n=427) (vgl. Krähwinkel 2015, S. 9) ist die Nachfrage nach Unterstützung bei den Themen Metadaten und Langzeitarchivierung mit 66,1 % sogar der meist gewünschte Service (vgl. S. 39). Interessanterweise steht dieses Angebot bei den

Gruppen Medizin und Naturwissenschaften auf Platz eins der Wünsche (vgl. S. 39).

Auch Wiley und Burnette von der *University of Illinois* haben in deren Untersuchung der FDM Praxis in *Bioengineering* und biomedizinischer Forschung beschlossen, dass unter den Forschenden bemerkenswerte Lücken bzgl. Kenntnissen und Bewusstsein für Metadaten und Dokumentation herrschen (vgl. 2019, S. e1132). Dies fordert „*instruction in the areas of file organization [...] documentation, and metadata standards*“ durch wissenschaftliche Bibliotheken (Wiley und Burnette 2019, S. e1132).

Die Ergebnisse der empirischen Untersuchungen weisen vor allem ein fehlendes oder mangelndes Bewusstsein und Kenntnisse im Bereich Metadaten vor. Dabei ist ein Bedarf an Informationsvermittlung und Beratung im Bereich der Beschreibung von Forschungsdaten zu sehen. Weiter zeigt sich der Wunsch nach möglichst einfachen, nicht aufwendigen Lösungen. An der vergleichsweise niedrigen Anzahl an Studien, die Metadaten bei der Bedarfserhebung thematisieren, stellt sich das Thema gegebenenfalls als nicht besonders relevant für die durchführenden Servicestellen dar.

3 Methodik

Das Ziel der vorliegenden Arbeit ist, die praktischen Ansätze in der Beschreibung von Forschungsdaten, sowie die Bedürfnisse und Anforderungen der Forschenden an Services in diesem Bereich darzustellen. Wie aus der Literaturübersicht ersichtlich, sind die Begriffe „Forschungsdaten“ und „Metadaten“ nicht unmittelbar eindeutig und werden in verschiedenen Disziplinen nicht einheitlich verstanden. Aus diesen Gründen wurde auf die Fragestellung mithilfe qualitativer Forschung eingegangen. Im Vergleich mit quantitativen Methoden, die sich auf das Testen von a priori formulierten Hypothesen konzentrieren, ist der Grundgedanke qualitativer Methoden „Erkenntnisse *bottom-up* aus dem empirischen Material zu gewinnen und ihrem Gegenstand mit größtmöglicher Offenheit zu begegnen“ (Umlauf, Fühles-Ubach und Seadle 2013, S. 65). Um die Frage nach Bedürfnissen und Anforderungen zu beantworten, scheint deshalb der Einsatz qualitativer Methoden als sinnvoll. Die Arbeit wurde in Anlehnung auf die *Grounded Theory* durchgeführt.

3.1 Grounded Theory

Die *Grounded Theory* wurde in den 60er Jahren des letzten Jahrhunderts von den US-amerikanischen Soziologen Barney G. Glaser und Anselm L. Strauss entwickelt als Reaktion auf die damalige Orientierung der Disziplin auf das Überprüfen von Hypothesen (vgl. 1967, S. 1-2). Im Gegensatz zu diesem Vorgehen, sollte die *Grounded Theory* „*the discovery of theory from data systematically obtained from social research*“ erlauben (Glaser und Strauss 1967, S. 2). Die Theorie sollte in diesem Sinne mit Daten nicht nur beispielhaft unterstützt werden, sondern aus diesen stammen, und in diesen „*grounded*“ sein (Glaser und Strauss 1967, S. 33). Wie Strübing erklärt, setzt der Grundgedanke voraus, dass Forschende möglichst aufgeschlossen und ohne vorherige theoretische Festlegungen an das Untersuchungsfeld herangehen (vgl. 2014, S. 59). Auch dies unterstützt die Entscheidung, einen Ansatz der *Grounded Theory* zu verwenden, um die Wünsche der Naturwissenschaftler_innen möglichst realitätsnah darzustellen. „In seiner Doppeldeutigkeit verweist die Bezeichnung *Grounded Theory* gleichermaßen auf Prozess und Ergebnis“ (Strübing 2013, S.

111). So wird bereits im Begriff selbst ein bedeutender Aspekt der Methodik vermittelt, nämlich sowohl die Theorieentwicklung als Prozess, als auch die am Ende dargestellte Theorie.

Seit der ersten Veröffentlichung von *The Discovery of Grounded Theory*, wurde diese in unterschiedliche Richtungen weiterentwickelt und findet Anwendungen weit über die Grenzen der Soziologie hinaus, für jene sie ursprünglich konzipiert wurde. Die vorliegende Arbeit orientiert sich dabei an den wesentlichen Punkten der *Grounded Theory* (vgl. Glaser und Strauss 1967; Glaser 1978; Strauss 1987), die sich folgendermaßen zusammenfassen lassen:

1. gleichzeitige Erhebung und Auswertung von Daten
2. Erstellung von analytischen Codes und Kategorien aus Daten, nicht aus im Voraus überlegten Hypothesen: Im Gegensatz zu der Art von Hypothesen, die zum Testen bereits bestehender Theorien verwendet werden, produziert der *grounded theorist* Hypothesen aus empirischen Daten, die von anderen getestet werden können.
3. Anwendung der *Constant Comparative Method*, bei der Vergleiche in jeder Phase der Analyse durchgeführt werden.
4. Die Theorie wird in jedem Schritt der Datenerhebung und -auswertung weiterentwickelt.
5. Schreiben von Memos, um Kategorien auszuarbeiten, deren Eigenschaften festzulegen, die Beziehung zwischen Kategorien zu definieren und Lücken zu identifizieren.
6. Das *Sampling* zielt nicht auf Repräsentativität der Bevölkerung ab, sondern auf die Erstellung einer Theorie.
7. Die Verfassung eines Literaturberichts findet erst nach der Erarbeitung einer unabhängigen Analyse statt.

(vgl. Zusammenfassung nach Charmaz 2014, S. 7-8; Bryant 2017, S. 90)

In der Arbeit kommen die Punkte 1.-5., sowie 7. erfolgreich zur Anwendung. Das *Theoretical Sampling* konnte bedauerlicherweise nicht durchgeführt werden, da

dies über den zeitlichen Rahmen der vorliegenden Arbeit hinausging. Auf diese Problematik wird im Abschnitt 3.3 nochmals im Detail eingegangen.

3.2 Das Leitfaden-Interview

Einer der Grundprinzipien qualitativer Forschungsmethoden zielt darauf hin „die Frage nach dem *Wie*“ zu beantworten (Strübing 2013, S. 24). Für die Datenerhebung wurde das teilstrukturierte Leitfadeninterview ausgewählt. Dies bietet sich als besonders geeignet an, da es die notwendige Offenheit zulässt, um den Forschenden die freie Darstellung der Praxis und der Vielfalt der Ansätze im Bereich Beschreibung von Forschungsdaten zu ermöglichen (vgl. Helfferich 2011, S. 180). Dies entspricht ebenso den Grundsätzen der *Grounded Theory*. Darüber hinaus erlaubt das Leitfadeninterview auch „stärker auf inhaltliche Informationen“ gerichtete Fragen zu formulieren (Helfferich 2019, S. 673). Dies wird für eine so konkrete Problematik, die u.a. Unterschiede zwischen Disziplinen ausarbeiten möchte, als relevant betrachtet.

Für die Interviews wurde anhand der SPSS-Methode (vgl. Helfferich 2011, S. 182-9) ein Leitfaden entwickelt. Eine Kopie ist im Anhang I zu finden. Im Leitfaden wurden obligatorische und weiterführende Fragen ausformuliert. Der gesamte Leitfaden sollte allerdings nicht als ein statisches Objekt betrachtet werden. Es handelt sich stattdessen um eine Zusammenführung möglicher Themenbereiche, die in unterschiedlicher Reihenfolge, anders als im Leitfaden formuliert oder, im Falle der weiterführenden Fragen, gar nicht im Gespräch gestellt werden müssen. Als Hauptproblem bei der Erstellung des Leitfadens hat sich der Umgang mit dem zentralen Begriff „Metadaten“ erwiesen. Da die Bekanntheit des Begriffs einen wichtigen Teil der Fragestellung ausmacht, wurde entschieden, neben dem Wort „Metadaten“ mehrere quasi synonyme Bezeichnungen wie „Beschreibung von Forschungsdaten“ und „Dokumentation“ im Gespräch zu verwenden. Um der eventuellen Unverständlichkeit des Begriffes weiter entgegenzuwirken, wurde eine Definition vorbereitet, die bei Nachfragen zur Erklärung vorgelegt werden konnte. Es wurde diesbezüglich eine Kombination aus den Definitionen der Plattform forschungsdaten.org und der TU Wien Webseite zum selben Thema gewählt:

„Metadaten bezeichnen alle zusätzlichen Informationen, die zur Interpretation der eigentlichen Daten, z.B. Forschungsdaten notwendig oder sinnvoll sind und die eine (automatische) Verarbeitung der Forschungsdaten durch technische Systeme ermöglichen. Metadaten werden daher oft als 'Daten über Daten' bezeichnet und dienen dazu, die unterschiedlichen Informationen zu digitalen Objekten zu kategorisieren und zu charakterisieren“ (Helmholtz-Zentrum Potsdam, Deutsches GeoForschungsZentrum GFZ, o.D.).

„Metadaten dienen in Archiven zur einheitlichen und strukturierten Beschreibung der verwalteten Ressourcen und ihrer Merkmale. Sie sind an sich unabhängig und werden entweder gemeinsam mit den Ressourcen, die sie beschreiben, oder eigenständig abgespeichert. Metadaten spielen eine entscheidende Rolle für die zukünftige Nutzung Ihrer abgelegten Forschungsdaten: Sie enthalten wichtige Informationen und ermöglichen das Suchen und Finden Ihrer Daten“ (Technische Universität Wien 2019).

Der Begriff Metadaten schien mithilfe dieser Quellen als anschaulich erklärt und auch für „Laien“ verständlich zu sein, da sich beide im Vergleich mit wissenschaftlicher Literatur zum Thema vor allem an die Praxis der Forschenden orientieren.

Die Fragen im Leitfaden wurden in fünf Bereiche gegliedert. Die erste Frage erlaubt den Forschenden, sich und deren Forschungsgebiet kurz vorzustellen. Der nächste Fragenbereich sollte einen möglichst nahtlosen Übergang zum Thema Metadaten erlauben. Die Teilnehmenden werden gebeten, die Begriffe „Forschungsdaten“ und „Metadaten“ zu definieren. Falls notwendig, kommt an dieser Stelle die bereits erwähnte Definition zum Einsatz. Im Teil drei wird die Praxis in der Beschreibung von Forschungsdaten erfragt. So wird u.a. nach den möglichen Rollen, die Metadaten bei *Data Sharing* und Verwendung von Fremddaten spielen, gefragt. Da das zentrale Thema der Arbeit, falls nicht bereits aus der Praxis bekannt, nicht leicht zu verstehen ist, sollten die beiden Fragen den Forschenden erlauben, das schwer greifbare Konzept in Verbindung mit Beispielen aus der eigenen Praxis zu setzen und über diese frei sprechen zu können. Die vertiefenden Fragen umfassen weitere konkrete Themen wie die Sprache, in der Metadaten erstellt werden, oder den Bereich Schulungen und

Beratung im Bereich Forschungsdatenmanagement. Im letzten Abschnitt wird eine Frage nach den Wünschen der Teilnehmenden in Bezug auf Services zu Metadaten und eine offene Abschlussfrage eingeführt. Diese sollte dazu dienen, „dass die Befragten von sich aus in das Interview hineinbringen, was ihnen in diesem Zusammenhang wichtig ist“ (Krotz 2018, S. 61).

3.3 *Sampling* und Feldzugang

Wie bereits in der Literaturübersicht erwähnt, wurde diese Arbeit im Rahmen des Projektes e-Infrastructures Austria Plus durchgeführt. Nach Absprache mit der Leiterin des Arbeitspakets „Metadaten“ wurde entschieden, an erster Stelle die Personen zu kontaktieren, die folgende Bedingungen erfüllen:

- 1) naturwissenschaftliche Forschungstätigkeit
- 2) Interesse am Thema Forschungsdatenmanagement (z.B. vorherige Anfragen an die Datenmanagement-Abteilung der Universität Wien oder Teilnahme an Schulungsangeboten)

Der erste Kontakt erfolgte entweder per E-Mail, telefonisch oder persönlich nach der Teilnahme an einem Kurs in der Bibliothek. Es wurde versucht über das *Convenience Sampling* eine möglichst heterogene Gruppe zu generieren bezüglich Seniorität und Erfahrung, nämlich ab dem Masterstudium bis hin zu Professoren und Professorinnen. Weiter wurde der Schwerpunkt darauf gelegt, verschiedene naturwissenschaftliche Disziplinen einzubeziehen. Auf diese Art und Weise konnten nach mehreren Wochen fünf Teilnehmer für die Studie gewonnen werden. Dieser Ansatz stellt allerdings keinen Anspruch auf Repräsentation.

Trotz der beträchtlichen Bemühungen auch über das *Snowball Sampling*, konnten keine weiteren Teilnehmenden von der Universität Wien für die Studie gewonnen werden. Um diesem Problem entgegenzuwirken, wurde die Arbeit auf die Medizinische Universität Wien erweitert. Da die medizinische Forschung als eine naturwissenschaftliche Disziplin bezeichnet werden kann und die MedUni Wien ebenso am Projekt e-Infrastructures aktiv teilgenommen hat, schien dies als eine nicht besonders problematische Lösung. Dort konnten über persönliche Kontakte drei weitere Teilnehmer gewonnen werden. Trotz der kontinuierlichen Versuche

konnten keine weiteren Teilnehmer_innen im Zeitrahmen der Masterarbeit gefunden werden.

Im „*Theoretical Sampling* [...], *the analyst jointly collects, codes and analyzes his data and decides what data to collect next and where to find them, in order to develop his theory*“ (Glaser und Strauss 1967, S. 45). Bryant bemerkt, dass auch bei der *Grounded Theory* das *Convenience* oder *Purposive Sampling* am Anfang eines Forschungsprojektes eingesetzt werden kann als erster Schritt des *Samplings*, falls mit dem *Theoretical Sampling* möglichst früh danach angefangen wird (vgl. 2017, S. 112). Es ist zu beachten, dass die vorliegende Arbeit in diesem wichtigen Punkt von der *Grounded Theory* abweicht. Obwohl der Prozess der simultanen Datenerhebung und -analyse durchgeführt wurde, hat sich leider das gezielte Vorgehen bei der Suche nach Teilnehmenden im Zeithorizont der Masterarbeit und wegen des komplizierten Zugangs zu Forschenden als nicht praktikabel erwiesen.

Teilnehmer	Disziplin	Seniorität
Wissenschaftler A	Medizin	Junior Researcher
Wissenschaftler B	Chemie	Senior Researcher
Wissenschaftler C	Biochemie	Senior Researcher
Wissenschaftler D	Biologie	Senior Researcher
Wissenschaftler E	Physik	Senior Researcher
Wissenschaftler F	Biologie	Senior Researcher
Wissenschaftler G	Medizin	Senior Researcher
Wissenschaftler H	Medizin	Junior Researcher

Abbildung 1: Darstellung der Teilnehmenden und Forschungsdisziplinen

Obwohl die *Grounded Theory* nicht auf Repräsentativität abzielt, ist die oben angeführte Fallauswahl wegen der Stichprobenmethode kritisch zu sehen. Alle Wissenschaftler, die für die Studie akquiriert werden konnten, sind männlich. Trotz

des gezielten Ansprechens von Forscherinnen, konnte keine für die Studie gewonnen werden. Die Teilnehmer sind alle Vollzeit berufstätig und wurden nach Erfahrung in zwei Gruppen (Junior Researcher, Senior Researcher) aufgeteilt. Obwohl verschiedene Altersgruppen vertreten sind, wäre ein größerer Anteil an Junior Researchers willkommen. Die verschiedenen Forschungsdisziplinen sind ungleich verteilt mit drei Teilnehmern, die in der klinischen medizinischen Forschung tätig sind. Da es sich hier aber um einen sehr breiten Bereich handelt, stellt sich dies als nicht besonders problematisch dar. Die Hälfte der Befragten kommt ursprünglich aus Österreich, die zweite Hälfte aus anderen europäischen Ländern. Hier wäre bestimmt die Einbeziehung von Forschenden aus nicht europäischen Ländern ebenso von Interesse. Alle Befragten wohnen im Großraum Wien und sind an der Universität Wien oder der Medizinischen Universität Wien berufstätig.

3.4 Pre-Test

Am 11. April 2019 wurde der *Pre-Test* für die Leitfaden-Interviews durchgeführt. Dieser fand mit einer Person statt, die naturwissenschaftliche Forschung betreibt, um die Bedingungen der Stichprobe zu erfüllen. Das Interview wurde mit einem Smartphone aufgenommen, was sich als ausreichend erwiesen hat, um es auch für die darauffolgenden Interviews zu verwenden. Der Leitfaden wurde nach dem *Pre-Test* wegen Verständlichkeit angepasst. In der ursprünglichen Version wurde das „kollaborative Arbeiten“ als Begriff verwendet, dies wurde nach dem *Pre-Test* auf „Arbeit im Team“ geändert, denn dies schien als eher allgemein verständlich.

3.5 Durchführung der Interviews

Die Interviews fanden von 24.04.-11.05.2019 statt. Alle acht Teilnehmende wurden persönlich *face-to-face* befragt, um auch die Mimik und Gestik nachvollziehen zu können. Die Interviews dauerten zwischen 25 und 40 Minuten und wurden alle mit einem Smartphone aufgezeichnet. All die Interviews wurden anhand des Leitfadens durchgeführt. In den ersten zwei Interviews ergaben sich während der gleichzeitig stattfindenden Datenerhebung und Auswertung der Daten Verständlichkeitsprobleme mit dem Begriff „Metadaten“. Trotz dessen

haben diese Interviews in Bezug auf die Fragestellung interessante Ergebnisse geliefert und wurden daher für die weitere Auswertung beibehalten.

3.6 Transkribierung

Die Interviews wurden mehrmals angehört und es wurden mithilfe der Software MAXQDA Teiltranskriptionen erstellt. Während der Transkribierung wurden die Daten pseudonymisiert. Es handelt sich um das erweiterte „einfache Transkriptionssystem“ (Dresing und Pehl 2016, S. 20). Es wurde wörtlich transkribiert, inklusive vorhandener Dialekte. Aufgrund der besseren Lesbarkeit wurden Wiederholungen von Wörtern ausgelassen, außer diese dienten der Verstärkung der Aussagen. Nonverbale Äußerungen wie beispielsweise das Lachen wurden in Klammern mittranskribiert, falls diese Einfluss auf den Inhalt der Aussagen hatten. Die verbalen Zeichen des Zuhörens, wie „ich verstehe“ wurden in der Regel nicht transkribiert.

3.7 Auswertung

Trotz der hier nacheinander aufgeführten Schritte, sollen diese, wie bereits erwähnt, nicht als einander folgende, sondern als gleichzeitig stattfindende Aktionen im Sinne der *Grounded Theory* betrachtet werden (vgl. Glaser und Strauss 1967, S. 43). Die Analyse kam daher simultan mit der Datenerhebung zum Einsatz. Für die Auswertung wurden die analytischen Memos, sowie das Kodieren verwendet.

3.7.1 Memos

Die Ziele, die die Erstellung von analytischen Memos verfolgt, können mithilfe folgender mnemotechnischer Stütze zusammengefasst werden: „*MEMO*: *Mapping research activities; Extracting meaning from the data; Maintaining momentum; Opening communication*“ (Birks, Chapman und Francis 2008, S. 70). Um den Überblick über die Daten und theoretische Überlegungen nicht zu verlieren, wurden im Sinne von *Grounded Theory* während des gesamten Forschungsprojektes analytische Memos geführt. Während der Interviews erfolgte dies händisch. Die Memos wurden anschließend in die MAXQDA Software

übertragen und dort wurden während der Analyse weitere erstellt. Das Schreiben analytischer Memos gibt einem den Freiraum, die eigenen Ideen in Bezug auf die Daten zu erkunden (vgl. Birks und Mills 2015, S. 40). Diese Dokumentation der eigenen Gedanken erlaubt die Nutzung der Ideen auch Wochen und Monate später und fördert dadurch die Entwicklung von Theorie.

3.7.2 Kodieren (*Coding*)

Das Kodieren bedeutet den Prozess der Markierung einzelner Wörter, Sätze oder Abschnitte vom Text mit passenden Eigenschaften (vgl. Saldana 2013, S. 4). Diese kann sich zum Beispiel an die Häufigkeit der Erwähnung eines bestimmten Themas halten oder Einzigartigkeit hervorheben. Laut Dey, „*coding is an activity designed to generate theory systematically rather than to accumulate evidence*“ (1999, S. 257). Das sorgfältige *Coding* versichert im Sinne der *Grounded Theory*, dass die entwickelte Theorie tatsächlich *bottom-up* aus den Daten entsteht, statt der Auswahl passender Beispielsätze, um eine bereits entwickelte Hypothese zu bestätigen und illustrieren. In der ersten Phase des offenen Kodierens (auch *initial coding*) wurden sowohl *In Vivo Codes*, die direkt die Sprache der Befragten benutzen (vgl. Charmaz 2014, S. 134), als auch *sociologically constructed Codes* (selbst entwickelt) verwendet (vgl. Brüsemeister 2008, S. 158). *In Vivo Codes* eignen sich besonders gut für *Grounded Theory* Studien (vgl. Saldana 2013, S. 84). Sie wurden womöglich bei sowohl den englisch- als auch den deutschsprachigen Interviewtranskripten eingesetzt, um am Anfang der Auswertung möglichst nahe an den Originalaussagen zu bleiben. Dies wurde händisch auf Papier an kleinen Textsegmenten durchgeführt. In weiterer Folge wurden diese *Codes* in die Software MAXQDA übertragen und nochmal bearbeitet. Die *In Vivo Codes*, die aus den drei englischsprachigen Interviews entstanden sind, wurden über Vergleiche mit den entsprechenden Kodierungen auf Deutsch zusammengeführt, um die weitere Analyse aller Interviews einheitlich auf Deutsch durchführen zu können. In der zweiten Phase wurden alle *Codes* auf Zusammenhänge und Beziehungen geprüft und im Rahmen des selektiven Kodierens (*focused Coding*) in passende Kategorien zusammengefasst. Das *focused Coding* wird eingesetzt, „*to synthesize, analyze, and conceptualize larger segments of data*“ (Charmaz 2014, S. 138). Mit Augenmerk auf die Fragestellung wurden die Transkripte mehrmals gelesen und der Kodierungsprozess wiederholt,

um möglichst viele relevante Kategorien und Konzepte aus den Daten ausarbeiten zu können. Das daraus entstandene Codesystem ist auf den Seiten 24 und 25 grafisch dargestellt. Es handelt sich um die Kategorien „Metadaten“ (Abb. 2) und „Bedürfnisse und Anforderungen“ (Abb. 3). Die Kategorien „Forschungsgebiet“ und „Forschungsdaten“ sind im Codebuch im Anhang II nachzusehen.

3.8 Ethik

Bei qualitativen Forschungsmethoden sind natürlich die Fragen des ethischen Vorgehens von höchster Relevanz. Es handelt sich in diesem Falle um eine Tonaufnahme und Transkription von Gesprächen, die durchaus persönliche Meinungen und kritische Äußerungen beinhalten können. Die Daten wurden während der Transkription pseudonymisiert, um die Identifizierung der Teilnehmenden möglichst zu erschweren. Da die Arbeit im Rahmen des Projektes e-Infrastructures an Wiener Universitäten stattfand, war eine Anonymisierung, die z.B. die komplette Entfernung aller „Hinweise auf den organisatorischen und regionalen Kontext der Studie“ (Hopf 2016, S. 201) voraussetzt, allerdings aufgrund der Art des Projektes nicht möglich. Vor dem Interview wurde jedem Befragten eine umfassende, der Datenschutzgrundverordnung entsprechende Einwilligungserklärung vorgelegt. Weiter wurden alle Fragen der Teilnehmer z.B. zur weiteren Verwendung der Daten beantwortet, sowie auf deren Rechte ausdrücklich hingewiesen. Zudem wurde auch die gewählte Methode der Pseudonymisierung den Teilnehmenden im Detail erklärt. Die unterzeichneten informierten Einwilligungen liegen für alle Gesprächspartner vor.

Abbildung 2: Codesystem Teil 1 – Kategorie „Metadaten“

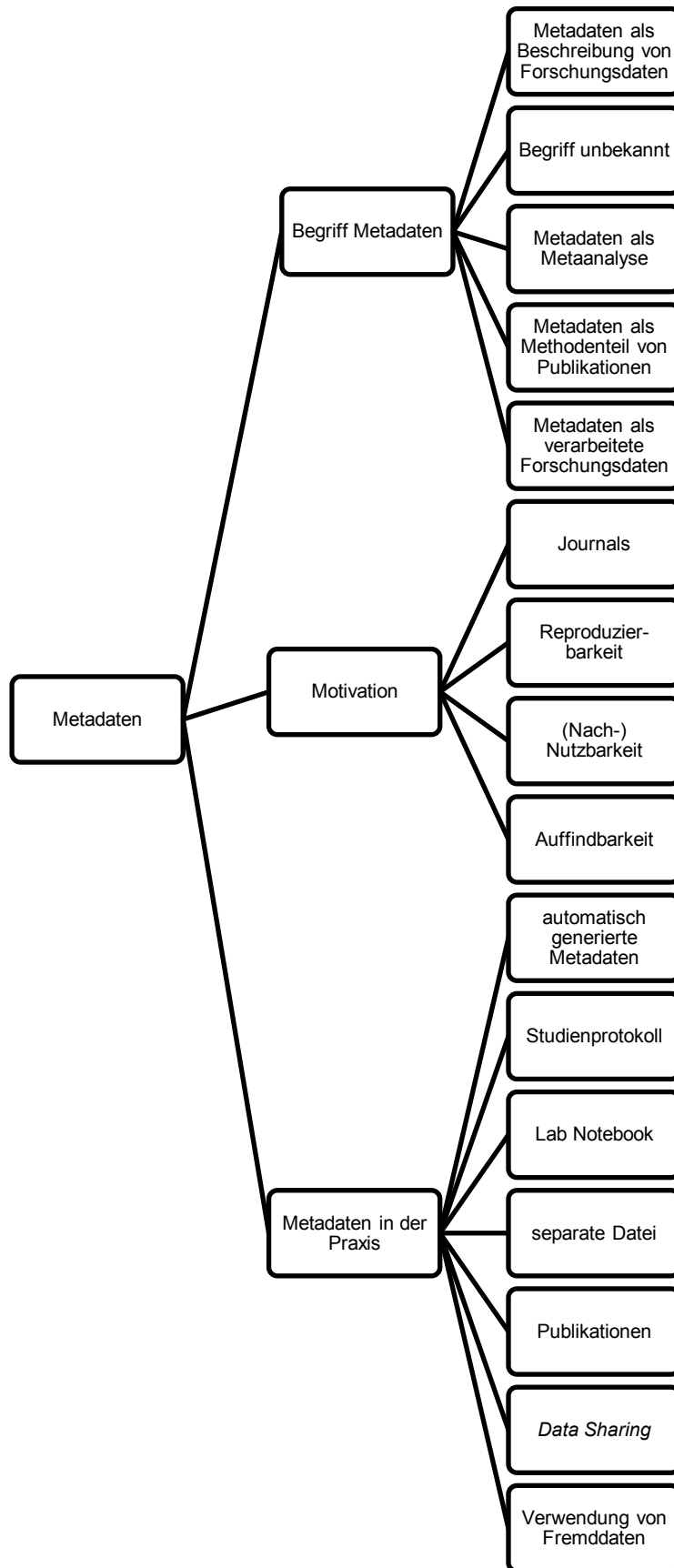
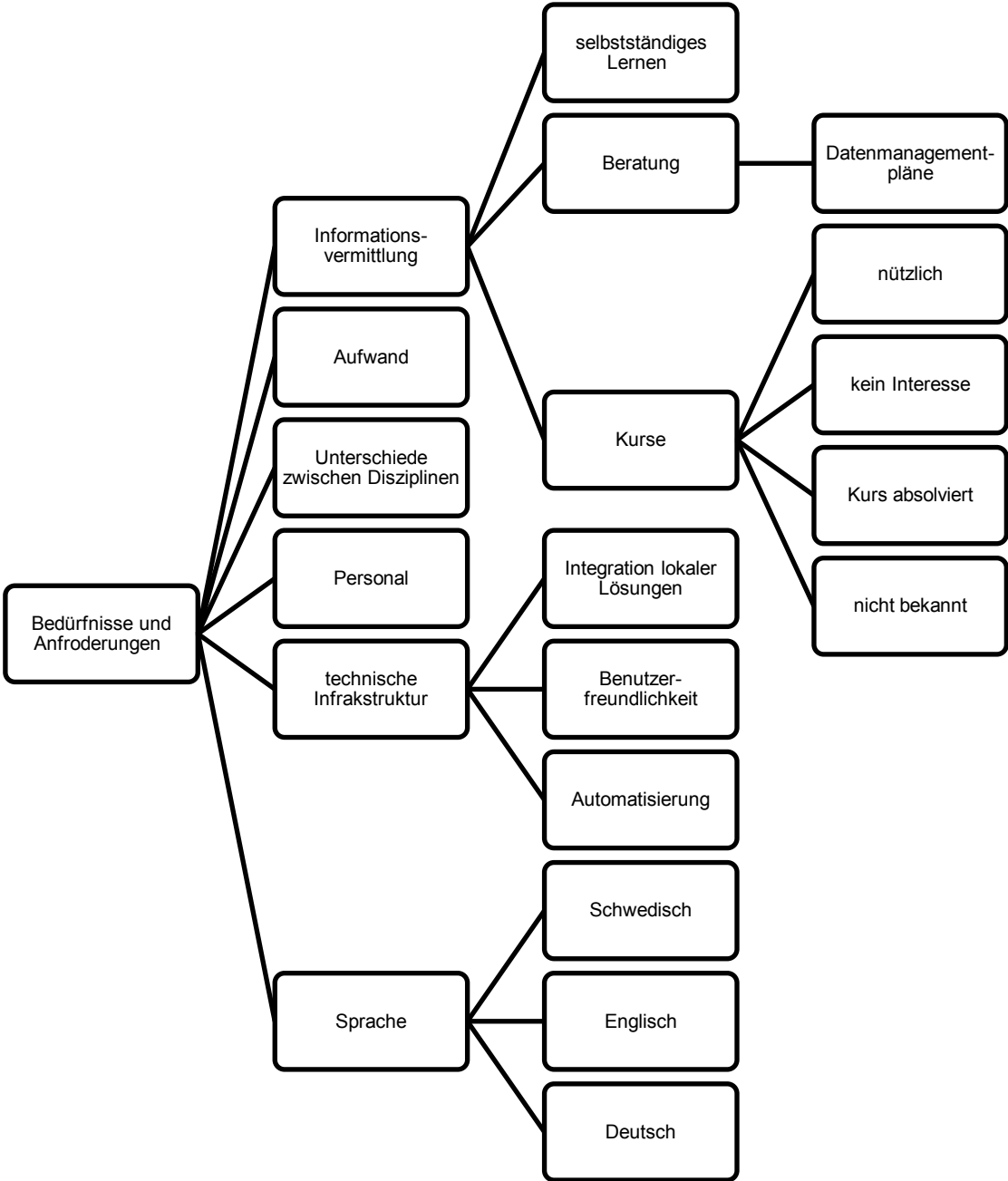


Abbildung 3: Codesystem Teil 2 – Kategorie „Bedürfnisse und Anforderungen“



4 Die Ergebnisse der qualitativen Interviews

4.1 Porträts der Wissenschaftler

In diesem Abschnitt werden die einzelnen befragten Wissenschaftler kurz skizziert. Demnach werden vor allem die Besonderheiten, die in den Interviews vorkamen, dargestellt.

Wissenschaftler A – Persönliches Interview, 24.04.2019

Der befragte Forscher hat letztes Jahr sein Medizinstudium abgeschlossen und befindet sich im ersten Arbeitsjahr an einer Universitätsklinik im Bereich der Inneren Medizin. Wie in der Medizin üblich ist, widmet er einen Teil seiner Arbeitszeit der klinischen Forschung als Junior Researcher. Er beschäftigt sich:

„konkret mit autoimmunen Muskelentzündungen. Das ist eine Krankheitsgruppe, die sehr selten ist. Ähm, sch/ unter Umständen auch schwer zu behandeln, aber es tut sich vieles“ (A. 3).

In diese Richtung baut er zurzeit ein s.g. Patientenregister, eine Datenbank, die alle Patienten der Abteilung mit einer konkreten Krankheit abbildet, auf. Solche Datenbanken werden in der klinischen Medizin sowohl zu der Erfassung von Patienten als auch zu der Generierung neuer Hypothesen eingesetzt (vgl. Workman, o.D.). Er spricht viel von Kooperation und Interdisziplinarität, die in seinem Bereich eine bedeutende Rolle spielen:

„[n]achdem die, natürlich, die ethischen Voraussetzungen geklärt sind ähm, wir kooperieren. Kooperation. ... Ich biete meine Daten an, bekomme dafür eine Ko-Autorenschaft. Weil ich natürlich im Gegenzug auch ein Projekt plane“ (A. 43).

Das Thema *Data Sharing* hält er in diesem Zusammenhang für sehr wichtig und plant das Patientenregister in Zukunft auch auf andere Krankenhäuser zu erweitern.

Wissenschaftler B – Persönliches Interview, 29.04.2019

Der Befragte kommt ursprünglich aus Österreich. Bereits seit vielen Jahren im Bereich organische Chemie wissenschaftlich tätig, befasst er sich in seiner Arbeit neben der Lehrtätigkeit mit der Erforschung von

„Interaktionen von Molekülen. Wie bilden sich, ähm, nicht kovalente [...] Bindungen zwischen Molekülen und was für Funktionen hat das Ganze zum einen in der Natur“ (A. 3).

Seine Arbeit ist durch Kooperation mit anderen naturwissenschaftlichen Disziplinen, wie der Physik und der Biologie im Rahmen der eigenen Universität, und auch durch die Teilnahme an großen nationalen und internationalen Projekten stark geprägt. Dementsprechend reflektiert er im Interview besonders intensiv die Unterschiede zwischen Disziplinen und die Schwierigkeiten vor allem in der Kommunikation bei gemeinsamen Forschungsprojekten, die sich daraus ergeben:

„man ist sich natürlich darüber klar, wir sprechen jeweils eine unterschiedliche Sprache. Also, Biologen reden anders als Chemiker und wir reden wieder anders als Physiker. Trotzdem müssen wir irgendwo miteinander dann klarkommen“ (A. 24).

Wissenschaftler C – Persönliches Interview, 29.04.2019

Der Befragte ist hauptsächlich in den Bereichen Zellbiologie und Biochemie tätig. Obwohl ursprünglich nicht aus Österreich arbeitet er nach vielen internationalen beruflichen Stationen seit mehreren Jahren als Senior Researcher in Wien. Das Interview macht deutlich, dass er sich stark mit den Themen des Forschungsdatenmanagements auseinandersetzt. Des Weiteren ist es ihm ein Anliegen, Bewusstsein für diese Themen inklusive Metadaten sowohl in seinem Team, als auch unter Studierenden zu schaffen:

„I try to really instill this in the lab that everyone has to not only provide the data but also to describe them“ (A. 26).

Er möchte Forschung mithilfe moderner Methoden betreiben, um international konkurrenzfähig zu bleiben. Effizienz und Innovation nehmen bei ihm einen

zentralen Platz ein. Er betont, dass sein Forschungsgebiet die Reproduzierbarkeit der Forschungsergebnisse fordert. Diesbezüglich betrachtet er die jetzige übliche Praxis kritisch:

„after I try to reproduce a thing and I write to people that we tried to do this way and it didn't work, they said, but of course because you used something different. This is something which is self-obvious for them but they did not really transfer this information to a publication or to a protocol“ (A. 35).

Wissenschaftler D – Persönliches Interview, 29.04.2019

Der Befragte ist Österreicher und ist seit vielen Jahren als Senior Researcher im Bereich Zoologie tätig. Dies findet in der Regel in einem kleinen Team in seiner Abteilung statt. In seinem Interview kommt zum Ausdruck, dass er neue Erkenntnisse verschiedener Forschungsgebiete für unterschiedlich bedeutend hält. Die wissenschaftlichen Ergebnisse in seiner Disziplin können beispielsweise nicht mit der Humanmedizin verglichen werden:

„[w]ir hamma keine Krebsforschung und ka Parkinson-Forschung, wo sich die ganze Welt auf a Thema stürzt und des von dem ganzen Globus hinweg sofort Aufsehen erregt, wenn jemand ähm was Neues findet“ (A. 15).

In Bezug darauf spricht er in seinem Gebiet von eher traditionellen Methoden der Publikation in wissenschaftlichen Zeitschriften. Die Ergebnisse in seinem Bereich werden als nicht relevant genug für die Gesellschaft betrachtet, um sich mit der Veröffentlichung von Forschungsdaten über den Rahmen eines publizierten Artikels zu beschäftigen. Er spricht diesbezüglich auch von der Wahrnehmung der Forschenden in den Medien, die seiner Meinung nach nicht der Realität entspricht:

„diese ähm ... Dinge, die durch die Medien gegangen sind, dass die Forschung so mogelt, und dass dort unseriöse Daten publiziert werden. [...] Das wird von den Medien verdreht. Weil es einfach besser klingt, wir brauchen ein Skandal“ (A. 19-20).

Wissenschaftler E – Persönliches Interview, 30.04.2019

Der Wissenschaftler ist als Senior Researcher im Bereich der Physik tätig. Ursprünglich nicht aus Österreich, arbeitet er seit mehreren Jahren wissenschaftlich in Wien. Er stellt sich als sehr informiert über die Themen des Forschungsdatenmanagements vor und betont, dass dies nicht zwingend das Niveau der Kenntnisse in diesem Bereich unter seiner Kolleg_innen widerspiegelt:

„[y]es, I know a fair bit about text and data mining and machine readability and interoperable metadata standards and repository interoperability and I've been following this Open Data space a bit. [...] Yeah, I expect that I have more knowledge than 90 percent of people on this topic“ (A. 12, 43).

Er thematisiert die Notwendigkeit, statt kommerzieller Software, Open Source und öffentlich finanzierte Lösungen zu verwenden. Seiner Meinung nach spielt Reproduzierbarkeit von Ergebnissen eine zunehmend bedeutende Rolle und er setzt sich dementsprechend für Open Science ein:

„having the raw data file available as Open Data, I think it improves the research product, the article itself. And that's the main form we concentrate on“ (A. 33).

In diesem Zusammenhang ist ihm die Modernisierung der Publikationsmöglichkeiten ein Anliegen.

Wissenschaftler F – Persönliches Interview, 09.05.2019

Der Befragte arbeitet als Senior Researcher im Bereich Biologie. Seine Forschungstätigkeit wird stark durch internationale Kooperationen geprägt:

„I am one of the most collaborative persons I know. I am currently working with more than 20 people across the world. From, I have collaborations with Oxford, Yale, German universities, Swiss universities, universities in Turkey, China, all around the world really“ (A. 18).

Er spricht von kulturellen Unterschieden zwischen seiner Heimat und Österreich, vor allem beim Thema Zeit und Effektivität. Obwohl er Kenntnisse im Bereich Forschungsdatenmanagement zeigt, liegt sein Fokus klar auf seiner Forschungstätigkeit im Vergleich mit eher „bürokratischen“ Aufgaben:

„[s]ome people like classifying things and have a certain ease with this kind of administrative way of thinking. I do not. I tried to keep it to the absolute minimum“ (A. 16).

Neben dem Thema reproduzierbare Ergebnisse, schildert er auch Trends sowie zukünftige Entwicklungsmöglichkeiten in seinem Gebiet vor allem im Bereich *Big Data*:

„[v]ery large datasets are the future. It's absolutely unavoidable. I mean the future, if one wants to be relevant in science“ (A. 55).

Wissenschaftler G – Persönliches Interview, 10.05.2019

Seit kurzem arbeitet der Befragte als Senior Researcher in der Inneren Medizin an einem Universitätskrankenhaus in Wien. Die Mehrheit seiner bisherigen Forschungstätigkeit fand allerdings in Schweden statt, wo er nach seinem Studium für seine Promotion tätig war. Als einziger in der Gruppe spricht er daher nicht von seiner jetzigen Situation, da er sich zurzeit, was sein Forschungsinteresse betrifft, neu orientiert. Die Nutzung von Daten, die von anderen erhoben wurden, ist ein wichtiges Thema in seiner Forschung. Er beschreibt z.B. Studien, die er in Schweden anhand von Daten aus einer nationalen Patientendatenbank bekommen hat:

„in Schweden ist das so, Patienten werden oft in riesengroße Register eingeschlossen. [...] da werden Patienten mit den, mit großen und wichtigen Erkrankungen automatisch in sogenannte Qualitätsregister eingeschlossen, relativ automatisch“ (A. 7).

Obwohl ihm die Begriffe des Forschungsdatenmanagements nicht sehr bekannt scheinen, zeigt er im Gespräch das Interesse daran und setzt sich diesbezüglich mit seinen bisherigen Berufserfahrungen kritisch auseinander. Er spricht von der teilweise mangelhaften Digitalisierung, die den Zugang zu wertvollen

Forschungsdaten erschwert. Er beschreibt die Verwendung von Fremddaten im Rahmen eines Forschungsprojektes:

„[e]s war eine Mischung aus Zettel und Excel-Files. Und wo ich, der Fehler war, z. B. was auch passiert, Übertragungsfehler, eh, dann muss ich auf die Zettel zurückgehen. Und wenn die Zettel nicht lesbar sind, muss man zurückgehen auf die Rohdaten“ (A. 36).

Wissenschaftler H – Persönliches Interview, 11.05.2019

Der zweite Junior Researcher im Sample ist aus seinem Heimatland vor einem Jahr nach Wien gezogen, um eine Stelle als Arzt und Junior Researcher an einer Universitätsklinik anzunehmen. Er beschäftigt sich im Bereich der Inneren Medizin neben seiner Doktorarbeit ähnlich zum Wissenschaftler A mit der Erstellung eines Patientenregisters. Obwohl er bis jetzt eher eigenständig statt in einem Team gearbeitet hat, sieht er das Teilen von Forschungsdaten als wichtig für die weitere Entwicklung seines jetzigen wissenschaftlichen Vorhabens:

„[w]as ich jetzt mache, ein Register, monozentrisch jetzt [an der Universität] beginne und im Idealfall auch dann in andere Institutionen weitertragen möchte oder in Kooperation mit anderen. Das heißt, der Aspekt der, des *Data Sharings* wird dann ja sehr wichtig sein“ (A. 34).

Im Interview nimmt die Verwendung von Fremddaten eine wichtige Rolle, denn seine Doktorarbeit basiert auf diesen. In Bezug darauf spricht er auch von Interoperabilität von verschiedenen Formaten. Das Interview verdeutlicht, dass sich der Wissenschaftler mit Themen des Forschungsdatenmanagements auseinandergesetzt hat. Er formuliert den markanten Unterschied zwischen klinischer und präklinischer Forschung und äußert den Wunsch nach Services, die diesen Unterschieden Rechnung tragen.

4.2 Forschungsdaten

Die Definition von Forschungsdaten unterscheidet sich stark zwischen den verschiedenen Disziplinen. Unter den acht befragten Wissenschaftlern findet sich eine große Vielfalt an Datentypen. In der klinischen medizinischen Forschung (Wissenschaftler A, G, H) handelt es sich überwiegend um personenbezogene und patientenbezogene Daten. Weiter wird u.a. von Messergebnissen, Daten von Simulationen oder Scandaten gesprochen.

Des Weiteren werden unterschiedliche Begriffe wie „*primary data*“, „Originaldaten“ oder „Rohdaten“ verwendet. Außerdem zeigen die Interviews, dass die Wissenschaftler zwischen Forschungsdaten nach Grad der Verarbeitung unterscheiden. Es wird allerdings deutlich, dass diese Begriffe ebenso nicht einheitlich verstanden werden.

„[o]ur lab books entries are the primary data, include primary data as well as the secondary processed data“ (Interview Wissenschaftler C 2019, A. 5)

„[z]um einen gibt es irgendwo einen synthetischen Bereich. Ähm, da hat man am Ende natürlich eine Verbindung vorliegen, die man in irgendeiner Form charakterisieren muss. Ich würde jetzt die Verbindung selber nicht als Daten interpretieren, sondern die aus den dabei [...] herauskommende Informationen sind dann im Prinzip unsere Rohdaten“ (Interview Wissenschaftler B 2019, A. 5).

Eine starke Abgrenzung zwischen Forschungsdaten und Publikationen gibt es zudem bei wenigstens zwei der Befragten (Wissenschaftler A, D) scheinbar nicht:

„natürlich auch letztendlich die Verarbeitung in Form von Publikationen. Das würd' ich als Forschungsergebnis, -daten bezeichnen“ (Interview Wissenschaftler D 2019, A. 5).

Auf die Frage danach, wo er Fremddaten suchen kann, antwortet Wissenschaftler A:

„[w]ir haben da unsere ... Datenbanken online, diese *PubMed* ... wo man Publikationen herausnimmt. Und, die einfach sozusagen studiert“
(Interview Wissenschaftler A 2019, A. 39).

Während die meisten Befragten überschaubare Massen an erhobenen Daten erwähnen,

„was ich hab' reicht locker, für eine, ein Terabyte a externe Festplatte, ja. Das alles, was ich an Publikationen in meinem Leben lang gemacht hab', ja. So viel ist das an Daten nicht“ (Interview Wissenschaftler D 2019, A. 22),

weisen drei auf den (zukünftigen) Trend, große Datenmengen zu erzeugen, hin:

„[v]ery large datasets are the future. It's absolutely unavoidable. I mean the future, if one wants to be relevant in science“ (Interview Wissenschaftler F 2019, A. 55).

4.3 Metadaten

In diesem Abschnitt werden die Ergebnisse zu den Fragen nach Bekanntheit des Begriffs „Metadaten“, sowie der Metadaten und Dokumentation zugrundeliegenden Konzepten in den Naturwissenschaften, sowie die jetzige Situation in der Praxis dargestellt.

4.3.1 Begriff „Metadaten“

Die Interviews weisen nach, dass die Hälfte der Befragten den Begriff Metadaten wenigstens zum Teil richtig einordnen kann:

„[m]etadata are basically the way to describe the common properties of the data“ (Interview Wissenschaftler C 2019, A. 9),

„metadata are the descriptors you give to the file about what is the provenance of the file“ (Interview Wissenschaftler E 2019, A. 8),

„a file describing, allowing us to use the data in a proper way, a descriptor kind of“ (Interview Wissenschaftler F 2019, A. 24),

„letztendlich, vom griechischen Ursprung des Wortes her, „Meta“, als über etwas. Also, irgendwie Daten die Daten beschreiben“ (Interview Wissenschaftler H 2019, A. 9).

In der anderen Hälfte ist der Begriff entweder unbekannt oder wird unbewusst für andere Ideen verwendet, beispielsweise für „Metaanalysen“:

„[Metadaten], das ist die zweite Ebene der Verarbeitung von Daten. Also, wenn ich zum Beispiel einen Review schreibe, wie ich’s vor paar Wochen veröffentlicht hab’, wo ich dann alle Arbeiten zu einem bestimmten Thema zusammenfasse“ (Interview Wissenschaftler D 2019, A. 7)

oder als der Methodenteil von Publikationen:

„[i]ch muss natürlich mitangeben, wie ich die Daten erzeugt habe, was für ein Gerät ich genutzt habe, ähm dafür was von der Konzentration ich vielleicht die Verbindungen da reingetan habe. All diese Dinge muss ich

in den Publikationen natürlich mitangeben“ (Interview Wissenschaftler B 2019, A. 19).

Der Begriff Metadaten aber auch die Umschreibung als „Beschreibung von Daten“ wird nicht eindeutig im Sinne der Metadaten und Dokumentation, als separat von den Forschungsdaten verstanden. Dies wird von drei der Forschenden als eine Ebene der Verarbeitung der Daten verstanden:

„wir [nehmen] diese gezackten Linien und holen uns da Daten heraus [...] und haben dann lange Listen von unterschiedlichen Zahlenkolonnen, ähm die halt am Ende das Ganze beschreiben“ (Interview Wissenschaftler B 2019, A. 11).

4.3.2 Beschreibung von Forschungsdaten in der Praxis

Trotz der Missverständnisse in Bezug auf den Begriff „Metadaten“, wie im Abschnitt 4.3.1 erläutert, konnten in den Interviews viele Beispiele von Metadaten und Dokumentation von Forschungsdaten unter diesen oder anderen Bezeichnungen im Arbeitsalltag der Befragten identifiziert werden. Da zwischen diesen beiden Konzepten von den befragten Wissenschaftlern nicht stark unterschieden wurde, werden sie in diesem Abschnitt unter „Beschreibung von Forschungsdaten“ zusammengefasst. Aufgrund der Heterogenität in den Antworten der Forschenden, wird in den folgenden Abschnitten sowohl auf Gemeinsamkeiten als auch auf die verschiedenen Ansätze eingegangen, um der Vielfalt in der Praxis Rechnung zu tragen. Es wird weiter auf die Themen Metadaten beim kollaborativen Arbeiten und bei der Nachnutzung von Forschungsdaten eingegangen. Alle acht befragten Wissenschaftler beschreiben deren Forschungsdaten in irgendeiner Form.

a) Motivation für die Beschreibung von Forschungsdaten

In den Interviews konnten direkt oder indirekt vier Hauptgründe für die Beschreibung von Forschungsdaten identifiziert werden. Bei fünf der acht Wissenschaftler spielen Auffindbarkeit und Sichtbarkeit eine große Rolle:

„[a]lso es war teilweise nicht auffindbar, wo überhaupt Daten vorhanden sind. In welcher Form sie vorhanden sind, sind das jetzt wie gesagt Zettel oder ... Files irgendwo ähm. Und da muss man natürlich auch von einer Person wissen, dass es das überhaupt gibt“ (Interview Wissenschaftler G 2019, A. 56),

„[w]as sicher a Problem ist, das halt viele Dinge, das ist das gesamte Problem der Literatur halt, vorhanden sind aber ganz ganz bescheiden zitiert werden, ja. Das ist sicher ein Grund, (unv.) Metadaten, [,,] irgendwo hätte“ (Interview Wissenschaftler D 2019, A. 16).

Für die Hälfte der Befragten machen Metadaten und Dokumentation Daten nutzbar für die Wissenschaftler selbst, sowie nachnutzbar, falls andere sie verwenden wollen:

„if you just collect the primary data and they are not annotated properly, this information is lost basically. It's very difficult to reuse it, if not impossible to reuse it“ (Interview Wissenschaftler C 2019, A. 17),

„[a]s a practicing scientist, metadata for us is that you need to, when you have a data file you need to somehow understand months and years later, what this data file is about and the metadata is the context that allows you to understand the file“ (Interview Wissenschaftler E 2019, A. 12).

Weiter ermöglicht für drei der acht Wissenschaftler die Beschreibung von Forschungsdaten die Reproduzierbarkeit der Ergebnisse:

„science requires reproducibility. So, when we scan something, when we do an experiment, we have to record a minimal number of parameters that would theoretically allow the experiment to be repeated“ (Interview Wissenschaftler F 2019, A. 34).

Zwei der Forschenden sprechen von Anforderungen von wissenschaftlichen Journals als wesentlichem Motivationsfaktor für die Beschreibung von Forschungsdaten:

„[w]e use this as evidence to support our claims, but we also need to provide the metainformation about how exactly this data were collected. Where does the sample come from. A lot of publications now actually require you to provide this information together with the data“ (Interview Wissenschaftler C 2019, A. 17).

Weitere drei berichten von der Veröffentlichung von Forschungsdaten als Voraussetzung für Publikationen. Dies bedeutet allerdings nicht zwingend die Verwendung ausreichender Metadaten.

b) automatisch-generierte Metadaten

Drei der Befragten erwähnen, dass automatisch erzeugte Metadaten in deren Forschungsgebiet von Belang sind. Diese werden in Bezug auf die automatischen Metadaten wie zum Beispiel das letzte Bearbeitungsdatum, die durch verschiedene Software-Lösungen generiert werden, erwähnt. Zugleich können Metadaten auch von den verwendeten Geräten erzeugt und gespeichert werden. Diese spielen nur zum Teil eine Rolle während der aktiven Forschungstätigkeit:

„[t]here is a lot of metadata that gets automatically saved about the state of the instrument. And sometimes we go back to the metadata and check some parameter later on. But typically that is getting automatically written, getting automatically archived but we never really look at it“ (Wissenschaftler E 2019, A. 14),

„was für mich immer wichtig ist, sind Daten zur oder Metadaten bezüglich der Erstellung und des Zeitpunkts und der Bearbeitung ... von Daten. Also das ist einfach der Chronologie-wegen, dass ich auch immer nachvollziehen kann, wann ich was gemacht habe“ (Interview Wissenschaftler H 2019, A. 11).

c) **disziplinspezifische Ansätze**

Das Konzept der Beschreibung von Forschungsdaten wird in den naturwissenschaftlichen Disziplinen unterschiedlich gehandhabt. In der klinischen medizinischen Forschung müssen vor dem Anfang jedes neuen Forschungsvorhabens detaillierte Anträge an die Ethikkommission der Institution gestellt werden. Diese beinhalten ein Studienprotokoll, ein Dokument, das u.a. den konkreten Umgang mit Forschungsdaten beschreibt:

„[w]ir müssen natürlich angeben, was wir ähm produzieren oder was wir entnehmen. Und wo das aufbewahrt wird [...] und wer eigentlich dafür verantwortlich ist“ (Interview Wissenschaftler G 2019, A. 31).

Die Protokolle werden laut der Aussagen in der Regel nicht mit den Forschungsdaten aufbewahrt oder generell weiterverwendet.

Die Befragten aus den Bereichen Physik und Biochemie erwähnen die Nutzung von Lab Books und Elektronischen Lab Notebooks, um Forschungsdaten zu beschreiben. Die Einträge in diesen werden sowohl als Forschungsdaten als auch als Metadaten gesehen, da dort die konkreten Prozesse und Tätigkeiten detailliert beschrieben werden:

„[t]he simplest metadata is what the person writes in his or her lab book“ (Interview Wissenschaftler C 2019, A. 26).

Die Wissenschaftler C und F berichten zusätzlich von der Nutzung selbsterstellter separater Dateien zur Beschreibung von Forschungsdaten:

„for scan data for example we always have a descriptive file, I guess that would count as metadata“ (Interview Wissenschaftler F 2019, A. 16),

„so we have the metadata attached to the primary data. Or we have to just describe them separately. Usually as an entry in a lab book or an entry in a Word file and then attach them to the primary data“ (Interview Wissenschaftler C 2019, A. 13).

d) **Publikationen**

Die Wissenschaftler B und D sprechen von Publikationen als Beschreibung von Forschungsdaten. Hier geht es im Prinzip um den Methodenteil von wissenschaftlichen Artikeln, wo laut den Angaben alle relevanten Informationen zu der Erzeugung der Daten dargestellt werden. Dies sollte die Nachvollziehbarkeit und Nachnutzung erlauben. Weitere separate Metadaten oder Dokumentation der Daten werden in diesem Zusammenhang als überflüssig betrachtet:

„wenn ich wissen möchte, wie ich’s damals gemacht hab, dann schau’ ich in die Publikation nach, in den Methoden nach. Weil das wäre für mich doppelte beschreibender Methoden, wozu?“ (Interview Wissenschaftler D 2019, A. 14)

e) **Metadaten und Kollaboration**

Beim kollaborativen Arbeiten spielen Metadaten laut Angaben aller Befragten mit einer Ausnahme (Wissenschaftler C, wo Metadaten teilweise automatisch mit Daten verknüpft sind), keine Rolle. Wenn ein Projekt im Team stattfindet, werden die Daten gemeinsam bearbeitet und ausgewertet. Die Vermittlung relevanter Informationen über die Daten erfolgt in der Regel über direkte Kommunikation, beispielsweise im Gespräch oder über E-Mail:

„[m]it den Daten macht man ja entsprechende Interpretationen. Und ähm da muss man natürlich dann sehr intensiv miteinander reden, weil ein Physiker macht andere Interpretationen als ein Chemiker“ (Interview Wissenschaftler B 2019, A. 36),

„Metadata for us is more about archiving. When you are actively collaborating, you’re looking at the files together, you are discussing them. Um you’re creating a figure for an article, metadata that doesn’t really come into it much“ (Interview Wissenschaftler E 2019, A. 14).

Es wurden keine Unterschiede in diesem Ansatz zwischen lokaler (Universität), nationaler oder internationaler Kooperation thematisiert.

f) **Metadaten bei Verwendung von Fremddaten**

Außer bei Wissenschaftler D, der erwähnt, dass die Verwendung von Fremddaten in seiner Disziplin gar nicht stattfindet, haben alle anderen befragten Forschenden bereits Erfahrung mit der Verwendung von Fremddaten gemacht. Die Interviews machen deutlich, dass dies vor allem in der klinischen medizinischen Forschung (Wissenschaftler A, G, H) eine besonders wichtige Rolle spielt, wo z.B. Daten aus großen Patientenregistern zur Analyse entnommen werden. Wissenschaftler A äußert diesbezüglich sogar, von Fremddaten „abhängig“ (A. 35) zu sein. In diesem Zusammenhang schildert der Wissenschaftler H Probleme mit der Verständlichkeit der Daten ohne ausreichende Beschreibung:

„bei der Klassifizierung von bestimmten Patientenmerkmalen, wo man sich dann halt eines numerischen Systems bedient hat [...], hat eine Legende gefehlt. [...] Die Daten waren dann im Datensatz nur codiert als eins, zwei, drei, vier, fünf und ich wusste jetzt nicht welcher Patient hat eine Chemotherapie bekommen“ (Interview Wissenschaftler H 2019, A. 21).

Von den Antworten lässt sich schließen, dass Beschreibung für Forschungsdaten bei Fremddatennutzung von Bedeutung ist. Die Angaben deuten darauf hin, dass die Auffindung von relevanten Fremddaten, sowie die Beschreibung von diesen in der Regel über Kontakte im Team erfolgt. In über der Hälfte der Fälle erfolgt dies auf einer informellen Art und Weise, ähnlich zu der Situation beim kollaborativen Arbeiten:

„[d]on't think I've ever actually looked at the metadata of a file from another collaborator. So, I trust that they send me the correct measurement and take, then tell me in an email this was measured in these and these conditions“ (Interview Wissenschaftler E 2019, A. 22),

„das war eine Mitforscherin, die gesagt hat, [...] dass die Daten existieren und die passen jetzt zu den aktuellen Daten, ob man die nicht anschaut. [...] Ich glaub', das war jetzt nicht irgendwie offen einsehbar auf einem Server oder sowas“ (Interview Wissenschaftler G 2019, A. 39).

4.4 Bedürfnisse und Anforderungen

Wie die verschiedenen Fachgebiete der befragten Forschenden versprechen, sind die Bedürfnisse und Anforderungen gleichermaßen vielfältig. Es werden in diesem Abschnitt möglichst diejenigen erwähnt, die sich auf die Erstellung von Metadaten beziehen. Da diese aber häufig von allgemeinerem Forschungsdatenmanagement nicht zu trennen sind, werden auch solche als Teil der folgenden Auflistung aufgeführt.

4.4.1 Unterschiede zwischen Disziplinen

Wie bereits in den Porträts der einzelnen Forschenden sowie im Abschnitt 4.2 zu Forschungsdaten geschildert, weisen die verschiedenen vertretenen naturwissenschaftlichen Fachrichtungen eine große Heterogenität zwischen den Disziplinen und Unterdisziplinen auf. Dies wird von drei der Befragten (Wissenschaftler B, E, H) auch explizit thematisiert:

„wir sprechen jeweils eine unterschiedliche Sprache. Also, Biologen reden anders als Chemiker, und wir reden wieder anders als Physiker. Trotzdem müssen wir irgendwo miteinander dann klarkommen. Ähm, und ja, jeder versucht sich irgendwo darauf einzulassen, dem anderen das [...] mit einem Rahmen zu erklären, was derjenige dann da [...] an eigenen Daten vorliegen hat“ (Interview Wissenschaftler B 2019, A. 24).

In diesem Zitat wird auch konkret die „unterschiedliche Sprache“ in den verschiedenen Disziplinen erwähnt. Es wird demnach der Wunsch angedeutet, Wege zu einer „einheitlichen Sprache“ zu finden. Dieses Element scheint nicht ausschließlich bei interdisziplinärer Forschungstätigkeit von Bedeutung zu sein.

4.4.2 Aufwand

Drei der Wissenschaftler (C, E, F), die sich laut deren Angaben mit Metadaten in der Praxis viel beschäftigen, sprechen ausdrücklich von der großen Anstrengung, die mit der Erstellung von Metadaten und Dokumentation für Forschungsdaten verbunden ist. Insbesondere wird dies als zu zeitaufwändig empfunden:

„the thing is that properly describing and annotating data is um difficult, it's very time consuming“ (Interview Wissenschaftler C 2019, A. 25).

Dies wird auch bei den Anmerkungen zur Verwendung des institutionellen Repositoriums der Universität deutlich:

„even just filling in the minimal metadata that's relevant that still takes for each data file [...], I mean it's not a long process maybe five minutes or three minutes [...], but for doing a lot more Open Data, it's still a bit too much“ (Interview Wissenschaftler E 2019, A. 28).

Obwohl die anderen Befragten nicht explizit von Zeitaufwand bei der Erstellung von Metadaten sprechen, kommt das Thema Effizienz auch in anderen Bereichen in deren Tätigkeit häufig vor. Es kann daher davon ausgegangen werden, dass dies für sie auch in Zusammenhang mit der Beschreibung von Daten von Bedeutung wäre.

4.4.3 Sprache

In sieben der acht Interviews wurde darauf hingewiesen, dass die Wissenschaft zweisprachig betrieben wird. In sechs Fällen heißt dies die Kombination Englisch-Deutsch, in einem handelt es sich um Schwedisch-Englisch. Ein Befragter erwähnt die Verwendung von nur einer Sprache (Englisch).

Die deutsche Sprache spielt eine Rolle im Arbeitsalltag überwiegend wegen der geografischen Notwendigkeit am Standort Österreich:

„Deutsch ist natürlich jetzt a mal [...] weil Österreich“ (Interview Wissenschaftler A 2019, A. 25).

Die Interviews deuten darauf hin, dass Deutsch als Wissenschaftssprache bis zu der Promotion eine Rolle spielt, während darüberhinausgehende wissenschaftliche Tätigkeiten vor allem in englischer Sprache stattfinden:

„ich glaub' bis auf meine Doktorarbeit vor ewigen Zeiten nie wieder was in Deutsch geschrieben“ (Interview Wissenschaftler D 2019, A. 25),

Laut der Befragten wird Englisch als *lingua franca* in der Alltagskommunikation im Team eingesetzt. Drei der Befragten sind nicht Muttersprachler von Deutsch. Aufgrund der Internationalisierung der Wissenschaft, kommen sowohl Studierende als auch Forschende aus dem Ausland häufig ohne bzw. mit nicht ausreichenden Deutschkenntnissen nach Österreich:

„we have some people who don't speak German at all in the group. Um very often have students or postdocs, who come from abroad and in the beginning they don't speak any German“ (Interview Wissenschaftler E 2019, A. 26),

„[s]obald einer da ist innerhalb der Gruppe, der die Landessprache [...] nicht spricht, dann wir Englisch gesprochen“ (Interview Wissenschaftler B 2019, A. 32).

Die englische Sprache wird nicht nur aufgrund der Kommunikation im Alltag präferiert, sondern auch wegen der überwiegend auf Englisch verfassten wissenschaftlichen Publikationen. Alle Befragten sind sich einig, dass Englisch die wichtigste Publikationssprache der Wissenschaft ist:

„[i]ch hab' meistens auf Englisch kodiert, einfach weil es leichter war [...] in der Präsentation, das ist dann eh auf Englisch, es wird nachher auf Englisch publiziert“ (Interview Wissenschaftler G 2019, A. 44),

„for someone working in science of nature [...], all the literature's in English“ (Interview Wissenschaftler F 2019, A. 32).

4.4.4 Informationsvermittlung

Die Hälfte der Befragten kennt bereits Beratungsangebote im Bereich FDM und diese werden von denen als wichtig betrachtet. Zum Beispiel der Wissenschaftler D bestätigt die Relevanz solcher Angebote, obwohl er sie zum jetzigen Zeitpunkt nicht benötigt:

„da seh' ich momentan keinen Bedarf. Weil ich, wenn ich das Problem hätte, würd' ich mich an [das institutionelle Repositorium] wenden, oder

die Leute da anschreiben oder anrufen. Die würden mir sicher weiterhelfen“ (2019, A. 31).

Zu den thematischen Bereichen der Beratung wurde besonders die Unterstützung bei der Erstellung von Datenmanagementplänen (Wissenschaftler E, H), sowie bei Archivierungsmöglichkeiten (Wissenschaftler D) genannt.

Die Einstellung der Befragten zum Thema Schulungen unterscheidet sich sehr. Solche Angebote scheinen weder besonderes bekannt zu sein noch wahrgenommen zu werden (sechs der acht Wissenschaftler). Dagegen halten zwei der Forschenden (C, E) Kurse für die Vermittlung von Forschungsdatenmanagementansätzen, inklusive Metadaten für sinnvoll:

„I think an introductory course would be useful. In particular things ... how you manage the data, how you annotate them in a reasonable way, what exactly are metadata, which ones are useful for others“ (Interview Wissenschaftler C 2019, A. 50).

Der Wissenschaftler C schlägt weiter vor, die Bedeutung von Metadaten und Forschungsdatenmanagement generell bereits im Studium vorzustellen, um das Bewusstsein für das Thema möglichst frühzeitig zu schaffen. Wissenschaftler C und E haben beide bereits an wenigstens einer Schulung zu diesen Themen teilgenommen. Auch in diesem Zusammenhang betont der Wissenschaftler E, wie wichtig Zeit und Effizienz für ihn sind:

„I did one course. Research data management at [a University in Vienna] or something like that was the title. It was a half-day training. [...] It was good. I think it was nicely compact. I've done some other trainings where it's like two days and I feel that the content would have been just half a day“ (2019, A. 36, 38).

Wissenschaftler H erwähnt ein für ihn nicht passendes Angebot an Kursen seitens der Universität:

„[ich] sehe ab und an, dass es Angebote gibt [...] Aber das sind dann eher Vortragsreihen oder sowas. Konkrete Schulungen irgendwie für

vielleicht Junior Scientists oder so ähnlich habe ich in der Form jetzt noch nicht gesehen“ (2019, A. 43).

Von dieser Aussage lässt sich ein gewisser Bedarf ableiten, an einer auf seine konkreten Bedürfnisse angepassten Schulung teilzunehmen. Außerdem zeigt sich ein wachsendes Interesse am Thema Beschreibung von Forschungsdaten im Gespräch mit zwei weiteren Befragten (Wissenschaftler B und G):

„das war ‘ne interessante Fragestellung. Also, ich hab’ nicht wirklich darüber nachgedacht aber jetzt im Nachhinein natürlich, dass wo ich mal da so zurückdenke [...], also dass das eigentlich eine relativ ganz gute Idee ist, das von Anfang an richtig zu machen“ (Interview Wissenschaftler G 2019, A. 73).

Zwei der Befragten geben zusätzlich an, sich das notwendige Wissen alleine anzueignen:

„[i]ch lerne alles selber, beziehungsweise tu’ ich auf Empfehlungen und Erfahrungen der anderen Kollegen“ (Interview Wissenschaftler A 2019, A. 45).

4.4.5 Personal

Aufbauend auf den Themen der markanten Unterschiede zwischen verschiedenen naturwissenschaftlichen Disziplinen, sowie zwischen den unterschiedlichen Forschungsdatentypen und eingesetzten Methoden, sprechen drei der Forschenden (C, E, G) von der Notwendigkeit Fachpersonal zu haben:

„if this has to be done properly, you actually need a person to do just that“ (Interview Wissenschaftler C 2019, A. 40).

Es wird die Anforderung an disziplinspezifisches Fachwissen hervorgehoben:

„I think it's probably true for most physics research, most natural science research is that the data is generated in so many different ways that are very specific and no generalist data management support person can be

familiar with all the technical nuances. So, I think it kind of has to be in-house“ (Interview Wissenschaftler E 2019, A. 40).

Der Wissenschaftler C schlägt diesbezüglich auch die Gründung von Arbeitsgruppen von Forschenden, die konkrete Lösungen und Best Practices im Bereich FDM erarbeiten könnten. Zwei weitere Forschende (A, B) wünschen sich mehr wissenschaftliches Personal für die Verarbeitung der Daten selbst. Auch dies unterstreicht die Beziehung zwischen Arbeit mit naturwissenschaftlichen (Meta-)Daten und Fachkompetenz.

4.4.6 Technische Infrastruktur

Es konnten in allen Interviews Anforderungen technischer Natur identifiziert werden. So wurde die Notwendigkeit geäußert, einen dienstlichen Computer zur Verfügung zu haben (Wissenschaftler A, D). Weiter sollten laut den Forschenden technische Lösungen für Metadaten und Dokumentation benutzerfreundlich und möglichst automatisiert sein, sowie die lokalen Lösungen reibungslos integrieren. Die konkreten Informationen, die dokumentiert werden müssen, um Ergebnisse nachvollziehbar und reproduzierbar zu machen, sowie Daten (nach-)nutzbar zu machen, sind in den verschiedenen Disziplinen sehr unterschiedlich. Deshalb ist bei technischer Infrastruktur auch eine gewisse Flexibilität gefragt.

a) Benutzerfreundlichkeit

In drei Gesprächen kommt das Thema Benutzerfreundlichkeit vor. Es handelt sich hauptsächlich um die Behebung des (zeitlichen) Aufwands beim Ausfüllen der Metadatenfelder im institutionellen Repositorium, das Wissenschaftler E „*a user interface issue*“ (2019, A. 28) nennt. Dies gilt aber auch allgemein bei technischen Lösungen für Metadaten, wie aus dem Zitat von Wissenschaftler C ersichtlich:

„[a]nother thing which is very important, develop some tools, which make it easy (lacht) because this is very hard“ (2019, A. 50)“.

b) Integration lokaler Lösungen und Automatisierung

Bei allen acht Befragten kann die Notwendigkeit einer Beschreibung von Forschungsdaten, ob in Form von Publikationen, als Einträge in Lab Books,

Studienprotokollen oder Metadaten-Files identifiziert werden. Wissenschaftler C und F schlagen vor, die bestehenden Metadaten in z.B. institutionelle technische Lösungen zu integrieren:

„[w]e have this, the parameters that are important to replicate the science. I think what would be useful, would be to ask us for our metadata files“ (Interview Wissenschaftler F 2019, A. 65),

„the problem there, of course is that yeah, you upload a subset of data but very often it's very difficult to upload the metadata“ (Interview Wissenschaftler C 2019, A. 37).

Zwei weitere Befragte (Wissenschaftler E und H) sprechen diesbezüglich von möglichst automatisierten Prozessen der Übernahme von Metadaten. Wissenschaftler E erklärt dies anhand der lokalen Software-Lösung und der Hindernisse, die sie für ein breiteres *Data Sharing* darstellt:

„the software, while it is open source, it's not in no way a standard software. So, the files that it writes are not usable, something that you would put as an Open Data. So when we want to share it with other researchers we have to convert it to another format and then write the metadata by hand basically“ (2019, A. 30).

Darüber hinaus äußern zwei der Wissenschaftler (C und E), die die Nutzung von *Electronic Lab Notebooks* als Insellösungen auf Instituts- oder Fakultätsebene angeben, den Wunsch nach Kostenübernahme dieser Software durch die Universität.

5 Diskussion

Obwohl es zahlreiche Publikationen gibt, die die Vorteile des nachhaltigen Forschungsdatenmanagements und der Verwendung von Metadaten für Forschungsdaten belegen, und geeignete Metadatenstandards für viele Fachgebiete bereits entwickelt wurden (vgl. Research Data Alliance. o.D.), zeigen die Interviews, dass der Begriff Metadaten von Forschenden nur teilweise verstanden wird. Fördergeber wie die Europäische Kommission oder der FWF in Österreich, sowie wissenschaftliche Bibliotheken verwenden in deren Kommunikation und Informationsmaterialien, die sich an Forschende richten, üblicherweise die Wörter „Metadaten“ und „Metadatenstandards“. Beispielsweise stellt die im April 2019 veröffentlichte Datenmanagementplan-Vorlage vom FWF Forschenden die Frage: „[w]elche Metadatenstandards (falls vorhanden) werden verwendet und warum?“ (FWF, S. 1). Die Interviews lassen anklingen, dass nur die Hälfte der Forschenden mit dem Begriff „Metadaten“ etwas anfangen kann. Es bietet sich daher die Frage, wie verständlich der Fachbegriff ohne konkrete Erklärung für Wissenschaftler_innen ist? Können diese das Feld „Metadaten“ im Datenmanagementplan ohne Unterstützung selbständig ausfüllen? Hierbei bestätigen die Interviews die Empfehlung von Read und Surkis, *„[i]t is important when approaching researchers to speak the language of research, not the language of libraries, for example, talk about describing data, not about Dublin Core or metadata“* (2015, S. 156). Interessanterweise handelt es sich um vier der fünf Wissenschaftler, die außerhalb Österreichs entweder studiert oder in der Forschung tätig waren, die den Begriff richtig einordnen können. Dies kann an dem stärkeren Bewusstsein für das Thema oder der weiterentwickelten Infrastruktur für FDM in anderen Ländern liegen.

Weiter deuten die Interviews aber darauf hin, dass auch die Umschreibung des Konzepts als „Beschreibung von Daten“, wie von Read und Surkis empfohlen (2015, S. 156) und z.B. in der e-Infrastructures Studie verwendet, unter Forschenden nicht eindeutig genug erscheint. Dies stellt die Ergebnisse der Untersuchungen an verschiedenen Universitäten, sowie das e-Infrastructures Projekt (Bauer et al. 2015), die Fragen nach Beschreibung von Daten stellen, infrage. Die Interviews zeigen, dass wir noch weit von einem flächendeckenden

Verständnis des Begriffes Metadaten entfernt sind. Die Verwendung von Standards und die umfassende Beschreibung von Daten hängt in erster Linie mit dem Verständnis des Begriffes zusammen. So empfiehlt sich aufbauend auf Read und Surkis, das Wort „*speak*“ (2015, S. 156) wahrzunehmen, mit Forschenden weiter ins Gespräch zu kommen, um die Vorteile der Beschreibung von Daten im bibliothekarischen Sinne zu vermitteln. Aus den Interviews ergibt sich so ein weiterer Handlungsbedarf für Bibliotheken in deren Rolle als Vermittler von Informationskompetenz.

Trotz der Unbekanntheit des Begriffes, konnten in den Interviews viele unterschiedliche Ansätze in der Beschreibung von Daten in den Naturwissenschaften identifiziert werden. Dies deckt sich mit den Erkenntnissen aus der Studie von Bauer et al. (2015). Sie sind in Form von automatisch-generierten Metadaten, *Lab Notebook* Einträgen, Studienprotokollen in der Medizin, sowie als separate, selbst erstellte Dateien zu finden. Darüber hinaus werden von manchen Wissenschaftlern auch Publikationen als Beschreibung von Daten betrachtet. So könnte der Ansatz von Chao verwendet werden, die diesbezüglich von „*methods metadata*“ (2015, S. 83) spricht. Sie untersucht Artikel im Bereich Bodenökologie, um aus dem Methodenteil für die Disziplin potenziell relevante Metadatenfelder auszuarbeiten (vgl. Chao 2015, S. 86). Außerdem herrscht in dieser Erkenntnis auch Potenzial für die Kommunikationsstrategien von Bibliotheken. Da Beschreibung von Forschungsdaten in den Naturwissenschaften bereits in irgendeiner Form stattfindet, scheint der Weg zu Metadaten und Dokumentation von Forschungsdaten im bibliothekswissenschaftlichen Sinne nicht abwegig. Bibliotheken könnten in deren Kommunikation mit Forschenden den Bezug zu Praxisbeispielen der Beschreibung von Forschungsdaten in diesen Formaten herstellen, diese dadurch greifbarer machen, sowie die zusätzlichen Vorteile der Metadaten und Dokumentation vermitteln. Weiter scheint die tief gehende wissenschaftliche Untersuchung von Beispielen dieser Formate der Beschreibung als sinnvoll, um Anforderungen und Bedürfnisse der Forschenden an Metadaten weiter auszuarbeiten.

Obwohl bei kollaborativen Projekten Metadaten laut den Interviews keine bedeutende Rolle spielen, da diese Informationen in der Regel direkt im Team vermittelt werden, scheinen sie bei der Verwendung von Fremddaten für

Forschende von Belang zu sein. Auch die Wissenschaftler, die sich mit Beschreibung von Forschungsdaten nicht sehr auskennen, würden laut deren Aussagen zur Fremddatenverwendung von strukturierten und umfassenden Informationen über diese Daten profitieren. Demnach wurden auch vier Motivationsfaktoren für die Erstellung von Metadaten und Dokumentation identifiziert. Diese machen Daten sichtbarer und leichter auffindbar. Weiter ermöglichen sie die (Nach)-Nutzbarkeit von Daten, sowie die Reproduzierbarkeit der Ergebnisse. Nicht zuletzt ermöglichen sie, den Anforderungen von wissenschaftlichen Journals zu entsprechen. Hier lässt sich ein wachsendes Bewusstsein für die Vorteile von Metadaten und Dokumentation erkennen, das strategisch angewandt werden kann, um die Erstellung von Metadaten und Dokumentation in den Naturwissenschaften voranzutreiben.

Aus den Interviews konnten sechs Bereiche an Bedürfnissen und Anforderungen in Bezug auf Metadaten und Services ausgearbeitet werden. Trotz der Unterteilung in sechs Bereiche, stehen die Wünsche der Forschenden miteinander in Beziehung und sollten nicht einzeln betrachtet werden. Es stellt sich als nicht sehr überraschend dar, dass die Forschung in naturwissenschaftlichen Disziplinen überwiegend in der englischen Sprache stattfindet. Die Interviews liefern damit eine klare Aussage auch für deutschsprachige Forschungseinrichtungen, die Forschungsdatenmanagement-Services anbieten wollen – sie sollen möglichst bilingual konzipiert werden. Weiter wird Sprache im übertragenen Sinne in den Interviews thematisiert. Hier kommen Unterschiede zwischen verschiedenen Disziplinen zur Sprache, die einen gewissen Bedarf an interdisziplinären Standards darstellen. So sollten die Besonderheiten der unterschiedlichen Fachgebiete bei der Entwicklung von Services möglichst mitbedacht werden.

Das Thema notwendiger Fachkenntnisse aus den jeweiligen Disziplinen wird von drei der Forschenden, die sich mit Metadaten in der Praxis viel beschäftigen, betont. Es wird diesbezüglich der Wunsch geäußert, Fachpersonal für das Thema FDM zu bekommen. Dies entspricht den neusten Entwicklungen in institutionellem FDM. Es werden Fachexperten, die über eine Ausbildung in einer relevanten Disziplin verfügen, als *Data Stewards* direkt auf Fakultäts- oder Institutebene

eingesetzt. Dieses Konzept wird beispielsweise an der TU Delft⁴ in den Niederlanden oder an dem österreichischen Repositorium AUSSDA⁵ für empirische Sozialforschung bereits angewandt. Dies bestätigt, die von Read und Surkis (2015) und Martin (2013) geäußerten Bedenken, die Forschende gegenüber Bibliotheken als Akteure im FDM haben können. Ein weiteres Hindernis bei der Erstellung von Metadaten ist der damit verbundene Zeitaufwand. Dieses Ergebnis entspricht ähnlichen Studien, die an anderen Universitäten durchgeführt wurden (Delasalle 2013, Lemaire et al. 2016). So ergibt sich aus den Interviews der Wunsch nach möglichst einfachen Lösungen, die den Forschenden das Leben erleichtern. Die Frage der Effizienz spielt sowohl bei technischer Infrastruktur als auch bei Schulungen eine Rolle.

Andere Aspekte, die bei dem Thema technische Lösungen von Bedeutung sind, sind Benutzerfreundlichkeit, Automatisierung und die Integration lokaler Lösungen. Drei der Forschenden äußern den Wunsch nach benutzerfreundlichen effizienten Lösungen. So scheinen Ansätze wie *Usability Tests* von Metadatenschemata in Repositorien als vernünftig (vgl. Petritsch 2017, S. 202). Zwei Wissenschaftler, die viele Metadaten produzieren, schlagen vor, diese in die zentralisierten technischen Lösungen, z.B. in ein Repositorium, automatisiert zu übernehmen.

Weitere Kommentare der Forschenden lassen sich in die Kategorie Informationsvermittlung zusammenfassen. Die Hälfte der Wissenschaftler betrachtet Beratung als eine wichtige Maßnahme. Dies spiegelt die Erkenntnisse anderer Studien von Feldsien-Sudhaus und Rajski (2016) und Krähwinkel (2015) wider. Zwei der Befragten haben bereits an Schulungen im Bereich FDM teilgenommen und bringen deren gute Erfahrungen zum Ausdruck. Außerdem äußern zwei Studienteilnehmer, dass sie sich selber informieren. Diese Ergebnisse betonen die Notwendigkeit verschiedener Wege der Informationsvermittlung. Diese decken sich größtenteils mit den Angeboten von Bibliotheken (vgl. Tenopir et al. 2017). Es zeigt sich aber auch Entwicklungspotenzial, gezielte praxisnahe fachspezifische Angebote zu konzipieren, um Forschende in deren wissenschaftlichen Vorhaben zu

⁴ Für die TU Delft siehe: <https://www.tudelft.nl/en/library/current-topics/research-data-management/research-data-management/data-stewardship/>

⁵ Für AUSSDA siehe: <https://aussda.at/ueber-aussda/team/>

unterstützen. In Kombination mit dem Wunsch nach Effizienz, scheint auch der Ausbau von Informationsvermittlung zum selbständigen Lernen als erstrebenswert.

6 Fazit

Die vorliegende Arbeit fand im Rahmen des Arbeitspakets „Metadaten“ des österreichweiten Projektes e-Infrastructures Austria Plus statt. Angelehnt an die *Grounded Theory* wurden von April bis Mai 2019 insgesamt acht Leitfaden-Interviews mit Forschenden von der Universität Wien und der Medizinischen Universität Wien durchgeführt. Die Befragten sind in den Fachgebieten Medizin, Biochemie, Biologie, Chemie sowie Physik wissenschaftlich tätig. Die Auswertung wurde mithilfe des Kodierens und analytischer Memos im Sinne der *Grounded Theory* durchgeführt. Es ist allerdings wichtig zu betonen, dass aufgrund des schweren Feldzugangs und dem begrenzten Zeitrahmen der Arbeit kein *Theoretical Sampling* zum Einsatz kam und die folgenden Ergebnisse sollen dementsprechend kritisch reflektiert werden.

Anhand der Interviews war das Ziel, die Fragen nach Bedürfnissen und Anforderungen im Bereich Metadaten für Forschungsdaten zu beantworten. Dabei wurden die Bereiche der Bekanntheit des Begriffs „Metadaten“ sowie die Darstellung der Praxis in Beschreibung von Daten mit untersucht.

Die Ergebnisse lassen darauf schließen, dass der Begriff „Metadaten“ nur zum Teil verständlich ist. Trotz dessen können unterschiedliche Beispiele der Beschreibung von Forschungsdaten in allen Interviews identifiziert werden. Es handelt sich u.a. um automatisch-generierte Metadaten, Einträge in *Lab Notebooks* oder die Ausarbeitung von Methoden in Publikationen. Als Motivationsfaktoren für die Erstellung von Metadaten und Dokumentation konnten die Auffindbarkeit, Sichtbarkeit und (Nach-)Nutzbarkeit von Daten, sowie Reproduzierbarkeit der Ergebnisse und Anforderungen von wissenschaftlichen Journals identifiziert werden.

Darüber hinaus konnten sechs Bereiche an Bedürfnissen und Anforderungen aus den Interviews abgeleitet werden:

- Angebote an Services sowie technische Lösungen sollen zweisprachig (Englisch/Deutsch) konzipiert werden.

- (Zeit-)Aufwand bei der Erstellung von Metadaten sollte durch geeignete Lösungen möglichst gering gehalten werden.
- Interdisziplinäre Unterschiede sollen bei der Entwicklung von Services mitbedacht werden.
- Es soll Personal mit Fachwissen in den konkreten Disziplinen als *Data Stewards* zur Verfügung gestellt werden.
- Technische Lösungen sollen benutzerfreundlich, möglichst automatisiert konzipiert werden, sowie lokale Lösungen integrieren.
- Informationsvermittlung zum Thema soll in Form von Beratung und Schulungen erfolgen, sowie das selbständige Lernen ermöglichen.

Diese Ergebnisse eröffnen Bibliotheken und weiteren Anbietern von Forschungsdatenmanagement viele Möglichkeiten, das Bewusstsein für Metadaten und Dokumentation von Forschungsdaten in den Naturwissenschaften weiter auszubauen. Laut der *Grounded Theory*: „*theory as process can be presented in publications as a momentary product, but it is written with the assumption that it is still developing*“ (Glaser und Strauss 1967, S. 32). Obwohl die vorliegende Arbeit interessante Ergebnisse liefert, stellen diese nur die ersten Einblicke in den Bedarf von Naturwissenschaftler_innen im Bereich Metadaten für Forschungsdaten dar. Bedürfnisse und Anforderungen ändern sich kontinuierlich und fordern daher weitere Untersuchungen, beispielsweise in Form von *Usability Tests* konkreter Lösungen. Weiter könnten auch qualitative Beobachtungen von Fokusgruppen, wo sich Forschende zum Thema untereinander austauschen können, von Interesse sein.

Abbildungsverzeichnis

Abbildung 1: Darstellung der Teilnehmenden und Forschungsdisziplinen.....	24
Abbildung 2: Codesystem Teil 1.....	29
Abbildung 3: Codesystem Teil 2.....	30

Literaturverzeichnis

- Bauer, B., Ferus, A., Gorraiz, J., Gründhammer, V., Gumpenberger, C., Maly, N., ... Steineder, C. (2015). *Forschende und ihre Daten (Ergebnisse einer österreichweiten Befragung: Report 2015)*. <https://phaidra.univie.ac.at/o:407513>
- Bibliotheksdienst. (2016). Forschungsdatenmanagement und Bibliotheken. *Bibliotheksdienst*, 50(7), 614–615. <https://doi.org/10.1515/bd-2016-0075>
- Birks, M., Chapman, Y., und Francis, K. (2008). Memoing in qualitative research. *Journal of Research in Nursing*, 13(1), 68–75. <https://doi.org/10.1177/1744987107081254>
- Birks, M., und Mills, J. (2015). *Grounded Theory: A Practical Guide (2. Aufl.)*. London, Großbritannien: SAGE Publications.
- Blumesberger, S. (2015). Die Welt der Metadaten im Universum von Repositorien. *Mitteilungen der VÖB*, 68(3/4), 515–528. <https://phaidra.univie.ac.at/view/o:422867>
- Brüsemeister, T. (2008). *Qualitative Forschung: Ein Überblick (2. Aufl.)*. Wiesbaden, Deutschland: VS Verlag für Sozialwissenschaften.
- Bryant, A. (2017). *Grounded Theory and Grounded Theorizing: Pragmatism in Research Practice*. <https://doi.org/10.1093/acprof:oso/9780199922604.001.0001>
- Büttner, S., Hobohm, H. C., und Müller, L. (2011). Research Data Management. In S. Büttner, H. C. Hobohm, & L. Müller (Hrsg.), *Handbuch Forschungsdatenmanagement* (S. 13–24). Bad Honnef, Deutschland: Bock + Herchen.
- Büttner, S., Rumpel, S., und Hobohm, H. C. (2011). Informationswissenschaftler im Forschungsdatenmanagement. In S. Büttner, H. C. Hobohm, & L. Müller (Hrsg.), *Handbuch Forschungsdatenmanagement* (S. 203–210). Bad Honnef, Deutschland: Bock + Herchen.
- Charmaz, K. (2014). *Constructing Grounded Theory (2. Aufl.)*. London, Großbritannien: SAGE Publications.

Chao, T. (2015). Mapping Methods Metadata for Research Data. *International Journal of Digital Curation*, 10(1), 82–94. <https://doi.org/10.2218/ijdc.v10i1.347>

Delasalle, J. (2013). Research Data Management at the University of Warwick: Recent Steps Towards a Joined-up Approach at a UK University. *Libreas. Library Ideas*, 2(23), 97–105. <https://doi.org/10.18452/9044>

Deutsches GeoForschungsZentrum – GFZ Helmholtz-Zentrum Potsdam. (o.D.). Forschungsdaten. Abgerufen 11. März, 2019, von <https://www.allianzinitiative.de/archiv/forschungsdaten/>

Deutsche Forschungsgemeinschaft. (2015). Leitlinie zum Umgang mit Forschungsdaten. Abgerufen 14. März, 2019, von https://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf

Dey, I. (1999). *Grounding Grounded Theory: Guidelines for Qualitative Inquiry*. San Diego, USA: Academic Press.

Dresing, T., und Pehl, T. (2016). *Praxisbuch: Interview, Transkription & Analyse* (6. Aufl.). Marburg, Deutschland: Eigenverlag.

European Commission. (o.D.). Data management - H2020 Online Manual. Abgerufen 12. Mai, 2019, von http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

Feldsien-Sudhaus, I., Rajski, B., und Technische Universität Hamburg, Universitätsbibliothek. (2016). *Digitale Forschungsdaten für die Zukunft sichern: Umfrage zum Umgang mit Forschungsdaten an der TU Hamburg (Auswertung)*. Abgerufen von <https://tore.tuhh.de/bitstream/11420/1329/4/TUHH-Forschungsdatenumfrage-2016-Auswertung.pdf>

Fingerle, B. (2015, 12. August). Figshare: Gebt uns allen Euren wissenschaftlichen Output! | ZBW MediaTalk. Abgerufen 14. April, 2019, von <https://www.zbw-mediatalk.eu/de/2015/08/figshare-gebt-uns-allen-euren-wissenschaftlichen-output/>

Fonds zur Förderung der wissenschaftlichen Forschung FWF. (Januar 2019).
FWF Datenmanagementplan (DMP) Vorlage - Guide. Abgerufen 20. April, 2019,
von
https://www.fwf.ac.at/fileadmin/files/Dokumente/Open_Access/FWF_DMPTemplate_d.pdf

Forschungsdaten.org – Universität Konstanz. (o.D.). Metadaten und
Metadatenstandards. Abgerufen 15. April, 2019, von
<https://www.forschungsdaten.info/themen/aufbereiten-und-veroeffentlichen/metadaten-und-metadatenstandards/>

Glaser, B. G. (1978): *Theoretical sensitivity: advances in the methodology of grounded theory*. Mill Valley, USA: Sociology Press.

Glaser, B. G., und Strauss, A. L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Hawthorne, USA: Aldine Publishing Company.

GO Fair. (2019). FAIR Principles - GO FAIR. Abgerufen 15. Mai, 2019, von
<https://www.go-fair.org/fair-principles/>

Helfferrich C. (2019) Leitfaden- und Experteninterviews. In: Baur N., Blasius J. (Hrsg) *Handbuch Methoden der empirischen Sozialforschung*. Wiesbaden, Deutschland: Springer VS.

Helfferrich, C. (2011). *Die Qualität qualitativer Daten: Manual für die Durchführung qualitativer Interviews* (4. Aufl.). Wiesbaden, Deutschland: VS Verlag für Sozialwissenschaften.

Helmholtz-Zentrum Potsdam, Deutsches GeoForschungsZentrum GFZ. (o.D.).
Metadaten. In [forschungsdaten.org](https://www.forschungsdaten.org). Abgerufen 8. April, 2019, von
<https://www.forschungsdaten.org/index.php/Metadaten>

Hey, A. J. G. und Trefethen, A. E. (2003). *The Data Deluge: An e-Science Perspective*. Wiley and Sons. Abgerufen 8. April, 2019, von
<https://eprints.soton.ac.uk/257648/>

Hey, A. J. G., Tansley, S., und Tolle, K. M. (2009). *The Fourth Paradigm: Data-intensive Scientific Discovery*. Abgerufen von <https://www.microsoft.com/en-us/research/publication/fourth-paradigm-data-intensive-scientific-discovery/>

Hopf C. (2016). Forschungsethik und qualitative Forschung. In: *Schriften zu Methodologie und Methoden qualitativer Sozialforschung*. Wiesbaden, Deutschland: Springer VS.

Jensen, U., Katsanidou, A., und Zenk-Möltgen, L. (2011). Metadaten und Standards. In S. Büttner, H. C. Hobohm, & L. Müller (Hrsg.), *Handbuch Forschungsdatenmanagement* (S. 83–100). Bad Honnef, Deutschland: Bock + Herchen.

Kindling, M., und Schirmbacher, P. (2013). „Die digitale Forschungswelt“ als Gegenstand der Forschung / Research on Digital Research / Recherche dans la domaine de la recherche numérique. *Information - Wissenschaft & Praxis*, 64(2-3). <https://doi.org/10.1515/iwp-2013-0017>

Krähwinkel, E. (2015). Forschungsdatenmanagement an der Philipps-Universität Marburg: Die Ergebnisse der Umfrage zum Forschungsdatenmanagement im November 2014. Abgerufen von <https://doi.org/10.17192/es2015.0019>

Krotz, F. (2018). Grounded Theory als integrierte Folge von Einzelfallstudien Besonderheiten eines Forschungsverfahrens. In C. Pentzold, A. Bischof und N. Heise (Hrsg.), *Praxis Grounded Theory: Theoriegenerierendes empirisches Forschen in medienbezogenen Lebenswelten. Ein Lehr- und Arbeitsbuch* (S. 53-72) Wiesbaden, Deutschland: Springer Fachmedien.

Lemaire, M., Rommelfanger, Y., Ludwig, J., Lürken-Uhl, A., Merkler, B., und Sturm, P. (2016). *Umgang mit Forschungsdaten und deren Archivierung. Bericht zur Online-Bedarfserhebung an der Universität Trier* (Universität Trier eSciences Working Papers, Nr. 2). Abgerufen von <https://ubt.opus.hbz-nrw.de/frontdoor/index/index/docId/737>

Letsch, M. (2017). *Aus- und Fortbildungsmöglichkeiten für das Forschungsdatenmanagement in Bibliotheken im deutschsprachigen Raum* (Masterarbeit). Abgerufen von <https://freidok.uni-freiburg.de/data/12867>

Ludwig, J., und Enke, H. (2013). *Leitfaden zum Forschungsdaten-Management: Handreichung aus dem WissGrid-Projekt*. Glückstadt, Deutschland: Werner Hülsbusch.

Martin, C. L. (2013). Wissenschaftliche Bibliotheken als Akteure im Forschungsdatenmanagement. *Libreas. Library Ideas*, 2(23), 12–21.
<https://doi.org/10.18452/9039>

Matusiak, K. K., und Sposito, F. A. (2017). Types of research data management services: An international perspective. *Proceedings of the Association for Information Science and Technology*, 54(1), 754–756.
<https://doi.org/10.1002/pr2.2017.14505401144>

Morgan, A., Duffield N. und Walkley Hall, L. (2017) Research Data Management Support: Sharing Our Experiences, *Journal of the Australian Library and Information Association*, 66(3), 299–305,
<https://doi.org/10.1080/24750158.2017.1371911>

Petritsch, B. (2017). Metadaten für Forschungsdaten in der Praxis. *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare*, 70(2), 200–207. <https://doi.org/10.31263/voebm.v70i2.1678>

Riley, J., und National Information Standards Organization U.S. (NISO). (2017). *Understanding Metadata: What is Metadata, and what is it For?*. Abgerufen 20. April, 2019, von
https://groups.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf

Rapport, F., und Braithwaite, J. (2018). Are we on the cusp of a fourth research paradigm? Predicting the future for a new approach to methods-use in medical and health services research. *BMC Medical Research Methodology*, 18(1).
<https://doi.org/10.1186/s12874-018-0597-4>

Research Data Alliance. (o.D.). Standards. Abgerufen 20. April, 2019, von
<http://rd-alliance.github.io/metadata-directory/standards/>

Saldana, J. (2013). *The Coding Manual for Qualitative Researchers* (2. Aufl.). London, Großbritannien: SAGE Publications.

Strauss, A. L. (1987). *Qualitative analysis for social scientists*. Cambridge, Großbritannien: Cambridge University Press.

Strübing, J. (2014). *Grounded Theory: Zur sozialtheoretischen und epistemologischen Fundierung eines pragmatistischen Forschungsstils* (3. Aufl.). Wiesbaden, Deutschland: VS Verlag für Sozialwissenschaften.

Strübing, J. (2013). *Qualitative Sozialforschung: Eine Einführung* (6. Aufl.). Abgerufen von <https://www.degruyter.com/view/product/224295>

Surkis, A., und Read, K. (2015). Research data management. *Journal of the Medical Library Association : JMLA*, 103(3), 154–156.
<https://doi.org/10.3163/1536-5050.103.3.011>

Technische Universität Wien. (2019). Metadaten. Abgerufen 8. April, 2019, von <https://www.tuwien.at/forschung/fti-support/forschungsdaten/aufbewahren-und-publizieren/metadaten/>

Tenopir, C., Talja, S., Horstmann, W., Late, E., Hughes, D., Pollock, D., ... Allard, S. (2017). Research Data Services in European Academic Research Libraries. *LIBER Quarterly*, 27(1), 23–44. <https://doi.org/10.18352/lq.10180>

Umlauf, K., Fühles-Ubach, S., und Seadle, M. (2013). *Handbuch Methoden der Bibliotheks- und Informationswissenschaft: Bibliotheks-, Benutzerforschung, Informationsanalyse*. Berlin/Boston: De Gruyter Saur.

Universität Innsbruck. (o.D.). Startseite. Abgerufen 27. Februar, 2019, von <https://www.e-infrastructures.at/de>

Universität Wien. (o.D.a). Zahlen, Daten & Publikationen. Abgerufen 20. April, 2019, von <https://www.univie.ac.at/ueber-uns/auf-einen-blick/zahlen-daten-broschueren/>

Universität Wien. (o.D.b). E-Infrastructures » Das Projekt. Abgerufen 14. April, 2019, von <https://e-infrastructures.univie.ac.at/das-projekt/>

Wiley, C. A., und Burnette, M. H. (2019). Assessing Data Management Support Needs of Bioengineering and Biomedical Research Faculty. *Journal of eScience Librarianship*, 8(1), e1132. <https://doi.org/10.7191/jeslib.2019.1132>

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1). <https://doi.org/10.1038/sdata.2016.18>

Wissenschaftler A (2019): Persönliches Interview, geführt von der Verfasserin am 24.04.2019.

Wissenschaftler B (2019): Persönliches Interview, geführt von der Verfasserin am 29.04.2019.

Wissenschaftler C (2019): Persönliches Interview, geführt von der Verfasserin am 29.04.2019.

Wissenschaftler D (2019): Persönliches Interview, geführt von der Verfasserin am 29.04.2019.

Wissenschaftler E (2019): Persönliches Interview, geführt von der Verfasserin am 30.04.2019.

Wissenschaftler F (2019): Persönliches Interview, geführt von der Verfasserin am 09.05.2019.

Wissenschaftler G (2019): Persönliches Interview, geführt von der Verfasserin am 10.05.2019.

Wissenschaftler H (2019): Persönliches Interview, geführt von der Verfasserin am 11.05.2019.

Workman, T. A. (o.D.). Defining Patient Registries and Research Networks - Engaging Patients in Information Sharing and Data Collection. Abgerufen 20. Mai, 2019, von <https://www.ncbi.nlm.nih.gov/books/NBK164514/>

Zook, M., Barocas, S., Boyd, D., Crawford, K., Keller, E., Gangadharan, S. P., ...
Pasquale, F. (2017). Ten simple rules for responsible big data research. *PLOS
Computational Biology*, 13(3), e1005399.
<https://doi.org/10.1371/journal.pcbi.1005399>

Anhang I – Leitfaden

[Persönliche Vorstellung, Einverständniserklärung, Technik]

Eröffnungsfrage

- Stellen Sie sich bitte vor und beschreiben Sie kurz Ihr Forschungsgebiet.

Offener Einstieg ins Thema

- Definieren Sie bitte, was in Ihrer Disziplin als Forschungsdaten gilt?
- Sagt Ihnen der Begriff Metadaten etwas? Was stellen Sie sich darunter vor? [bei Bedarf Definition vorlegen]

Metadaten in der Praxis

- Wenn Sie selber wissenschaftlich tätig sind, was für Metadaten erstellen Sie für Ihre Daten? Wie beschreiben Sie Ihre Daten?
- Schildern Sie bitte, wozu diese Metadaten dienen.
- Beschreiben Sie mir bitte, welche Rolle Metadaten spielen, wenn Sie im Team arbeiten (bspw. anhand des letzten Forschungsprojektes)?
- Wenn Sie Forschungsdaten von Anderen verwenden, welche Erfahrung haben Sie mit Metadaten von diesen Datensätzen gemacht?

Vertiefende Fragen

- In welcher Sprache beschreiben Sie Ihre Forschungsdaten?
- Wenden Sie Metadatenstandards an? Um welche handelt es sich?
- Welche Probleme hatten Sie möglicherweise bei der Beschreibung von Forschungsdaten?
- Was ist Ihnen bei den Metadaten wichtig, wenn Sie Forschungsdaten von Anderen benutzen?
- Stellen Sie sich vor, dass nach dem Abschluss Ihres Forschungsprojektes eine andere Person Ihre Daten nutzen möchte. Wie wäre dies möglich?
- Sind Ihnen Schulungs- oder Beratungsangebote bezüglich Forschungsdatenmanagement bekannt?

Abschluss

- Wenn alles möglich wäre, welche Services würden Sie sich von der Universität für die Erstellung von Metadaten wünschen?
- Wollen Sie zum Thema noch irgendetwas ergänzen?

[Bedanken für die Teilnahme; Ausfüllen von einem separaten Faktenfragebogen: z.B. Alter]

Anhang II – Codebuch

Forschungsgebiet

In dieser Kategorie werden die *Codes* zu den vertretenen Forschungsdisziplinen zusammengeführt.

Physik

Biologie

Biochemie

Chemie

Medizin

Forschungsdaten

Die Kategorie „Forschungsdaten“ beinhaltet alle *Codes*, die sich auf Forschungsdaten in den verschiedenen Disziplinen beziehen.

Forschungsdaten in Disziplinen – Welche Arten von Forschungsdaten werden in den jeweiligen Fachgebieten erhoben/erzeugt?

Publikation von FD – Wie werden Forschungsdaten publiziert?

Ebenen der Verarbeitung von FD – Welche Ebenen der Verarbeitung von Forschungsdaten gibt es?

Big data? – Welche Größe haben die erhobenen Daten?

Metadaten

Unter der Kategorie „Metadaten“ werden drei Unterkategorien zusammengeführt. „Begriff Metadaten“ umfasst alle Kodierungen der verschiedenen Definitionen von Metadaten. Im Bereich „Motivation“ finden sich die *Codes*, die als Motivationsfaktoren für die Erstellung von Metadaten identifiziert wurden. Bei der Unterkategorie „Metadaten in der Praxis“ werden alle *Codes* erwähnt, die praktischen Ansätze in der Beschreibung von Forschungsdaten darstellen. Außerdem sind hier weitere Aspekte, die in der Praxis eine Rolle spielen, zu finden.

Begriff Metadaten – Wie wird der Begriff Metadaten verstanden?

- Metadaten als Beschreibung von Forschungsdaten
- Begriff unbekannt
- Metadaten als Methodenteil von Publikationen
- Metadaten als Metaanalyse
- Metadaten als verarbeitete Forschungsdaten

Motivation

- Reproduzierbarkeit – Sind Metadaten wichtig, um Ergebnisse reproduzierbar zu machen?
- Journals – Haben Publikationsbedingungen von Journals einen Einfluss auf die Erstellung von Metadaten?
- (Nach-)Nutzbarkeit – Machen Metadaten Forschungsdaten nutzbar und nachnutzbar von anderen?
- Auffindbarkeit – Ermöglichen Metadaten die einfachere Auffindung von Forschungsdaten?

Metadaten in der Praxis

- automatisch-generierte Metadaten – Sind automatisch-generierte Metadaten wichtig?
- Studienprotokoll – Was sind Studienprotokolle und wofür werden sie verwendet?
- Lab Notebook – Wie werden (elektronische) Lab Notebooks benutzt?
- separate Datei – Wie werden separate Dateien als Metadaten eingesetzt?
- Publikationen – Was bedeutet Beschreibung von Forschungsdaten als Publikationen?
- *Data Sharing* – Spielen Metadaten eine Rolle beim *Data Sharing*?
- Verwendung von Fremddaten – Sind Metadaten bei der Verwendung von Fremddaten von Bedeutung?

Bedürfnisse und Anforderungen

Diese Kategorie fasst alle Kodierungen, die sich auf Bedürfnisse und Anforderungen beziehen, zusammen. Sie wird in sechs Bereiche gegliedert. Es handelt sich dabei um: „Informationsvermittlung“, „Aufwand“, „Unterschiede zwischen Disziplinen“, „Personal“, „Technische Infrastruktur“, sowie „Sprache“. Diese führen alle jeweils relevanten *Codes* zusammen.

Informationsvermittlung

- Beratung – Wie wird Beratung als Angebot gesehen?
 - DMP – Ist Beratung für Datenmanagementpläne erwünscht?
- Schulungen
 - nützlich – Wie werden Kurse betrachtet?
 - kein Interesse – Warum interessieren sich Forschende nicht für Schulungen?
 - Kurs absolviert – Welche Erfahrungen wurden mit Kursen gemacht?
 - nicht bekannt – Was ist der Bewusstseinsgrad bei Kursen?
- selbständiges Lernen – Wie können Forschende, die sich gerne selbstständig informieren und so lernen, unterstützt werden?

Aufwand – Welche Auswirkungen hat der (zeitliche) Aufwand auf die Erstellung von Metadaten?

Unterschiede zwischen Disziplinen – Welche Rolle spielen interdisziplinäre Unterschiede?

Personal – Wer soll für Metadaten und FDM zuständig sein und welche Eigenschaften benötigt das Personal?

- Arbeitsgruppen – Wie sollen Arbeitsgruppen für die Entwicklung von Standards eingesetzt werden?

Technische Infrastruktur

- Integration lokaler Lösungen – Wie können lokale Metadaten-Lösungen integriert werden?
- Benutzerfreundlichkeit – Wie können benutzerfreundliche Lösungen wie Metadatenfelder in Repositorien konzipiert werden?
- Automatisierung – Wie können Prozesse der Übernahme von Metadaten automatisiert werden?

Sprache

- Englisch – Wie kommt die englische Sprache zum Einsatz?
- Deutsch – Wie wird die deutsche Sprache verwendet?
- Schwedisch – Welche Rolle spielt die schwedische Sprache?

Anhang III – Transkripte der Interviews

Die hier aufgelisteten Transkripte sind öffentlich zugänglich und unter dem folgenden Link abrufbar: <https://doi.org/10.5281/zenodo.3897321>

- Wissenschaftler A – Transkript des Interviews vom 24.04.2019
- Wissenschaftler B – Transkript des Interviews vom 29.04.2019
- Wissenschaftler C – Transkript des Interviews vom 29.04.2019
- Wissenschaftler D – Transkript des Interviews vom 29.04.2019
- Wissenschaftler E – Transkript des Interviews vom 30.04.2019
- Wissenschaftler F – Transkript des Interviews vom 09.05.2019
- Wissenschaftler G – Transkript des Interviews vom 10.05.2019
- Wissenschaftler H – Transkript des Interviews vom 11.05.2019