

Publikationspraktiken für Forschungsdaten in Hochschulschriften

Eine Untersuchung der Veröffentlichungsformate und -methoden

eingereicht von

Dr. David Krassnig

als

Masterarbeit

im Rahmen des Weiterbildenden Masterstudiengangs

Bibliotheks- und Informationswissenschaft im Fernstudium

an der

Humboldt-Universität zu Berlin

Philosophische Fakultät


Institut für Bibliotheks- und Informationswissenschaft

1. Gutachter/in: Dr. Sarah Dellmann

2. Gutachter/in: Prof. Dr. Robert Jäschke

Berlin, 14.06.2024

Autor: Dr. David Krassnig

 <https://orcid.org/0000-0002-1626-7987>

Lizenz (Dokument): CC BY 4.0

DOI (Dokument): 10.5281/zenodo.11506621

Lizenz (Daten): CC0 1.0

DOI (Daten): 10.5281/zenodo.11401021

WASING THE ALWAYS OF WANTING OF KNOWING

–Brandon Sanderson, *Shadows of Self*



DIESE SEITE WURDE ABSICHTLICH LEER GELASSEN

Zusammenfassung

Diese Abschlussarbeit befasst sich mit den Forschungsdaten in Dissertationen und dem dazugehörigen Forschungsdatenmanagement an deutschen Universitäten. Die Arbeit lässt sich thematisch in zwei Teile gliedern.

Im ersten Teil werden allgemeingültige und promotionsspezifische Richtlinien einer Stichprobe aller deutschen promotionsberechtigten Institutionen auf Anweisungen zu Forschungsdaten untersucht. Dieser Bereich wird in Kapitel 3 behandelt.

Im zweiten Teil wird eine repräsentative, mehrschichtige Stichprobe aller Dissertationen aus dem institutionellen Repositorium der Leibniz Universität Hannover, die zwischen 2012 und 2023 erschienen sind, in Bezug auf Forschungsdaten untersucht. Hierbei steht besonders im Fokus, wie diese Forschungsdaten publiziert worden sind: als integrierter Bestandteil der Dissertation, als begleitende Dateien oder als separate Veröffentlichung auf einer externen Plattform. Dieser Bereich wird in Kapitel 4 behandelt.

Die Untersuchung der allgemeingültigen Richtlinien der promotionsberechtigten Institutionen ergab, dass etwa 45 % aller Institutionen in Deutschland eine allgemeine Forschungsdaten-Richtlinie besitzen (davon sind etwa 60 % Universitäten). Weitere 45 % besitzen zumindest eine Richtlinie zu den Regeln der Guten Wissenschaftlichen Praxis, während die restlichen 10 % keine allgemeine Richtlinie besitzen, die Forschungsdaten oder Forschungsdatenmanagement erwähnt (davon sind etwa 71 % künstlerische Hochschulen). Es wurde festgestellt, dass die Trägerschaft sowie die Institutionsart einen statistisch signifikanten Einfluss auf die höchste Klassifikationsstufe der allgemeinen Dokumente haben. Entsprechende mögliche Erklärungen zu den einzelnen signifikanten Relationen werden in Abschnitt 3.3.1 gegeben.

Die Untersuchung der promotionsspezifischen Richtlinien der promotionsberechtigten Institutionen ergab, dass nur etwa 14 % aller Institutionen in Deutschland mindestens eine Promotions- oder Prüfungsordnung besitzen, die den Umgang mit Forschungsdaten verpflichtend definiert. Von diesen Institutionen besitzen 75 % auch eine allgemeine Forschungsdaten-Richtlinie. Mehr als die Hälfte aller Institutionen haben jedoch zumindest eine Promotions- oder Prüfungsordnung, die verpflichtende Regularien zu den Regeln der Guten Wissenschaftlichen Praxis enthält. Ein weiteres Viertel der Institutionen besitzen keine Promotions- oder Prüfungsordnung, die Forschungsdaten oder Verpflichtungen zu den Regeln der Guten Wissenschaftlichen Praxis enthält. Es wurde festgestellt, dass das Bundesland, die Institutionsart einen statistisch signifikanten Einfluss auf die höchste Klassifikationsstufe der promotionsspezifischen Dokumente haben. Zusätzlich wurde festgestellt, dass die höchste Klassifikationsstufe der allgemeinen und promotionsspezifischen Richtlinien einer Institution signifikant voneinander abhängen. Entsprechende mögliche Erklärungen zu den einzelnen signifikanten Relationen werden in Abschnitt 3.3.2 gegeben.

Die Untersuchung der Dissertationen aus dem institutionellen Repositorium der Leibniz Universität Hannover ergab, dass etwa 87 % aller Dissertationen aus der

Stichprobe zumindest in Teilen auf originellen Primärdaten basieren. Von diesen Dissertationen haben etwa 61 % zumindest einen Teil ihrer Forschungsdaten publiziert. Dieser Anteil sinkt auf etwa 32 %, wenn komprimierte Fotografien, Spektraldiagramme, Gensequenzen, unausgefüllte Fragebögen, Leitfäden und Montagezeichnungen wegen mangelnder Qualität, Relevanz oder schwer zu evaluierender Originalität ausgeschlossen werden. Die häufigste Publikationsart waren in die Dissertationsdatei integrierte Forschungsdaten: Diese machten etwa 57 % aller Dissertationen mit originellen Primärdaten aus. Nur etwa 7 % aller Dissertationen mit Primärdaten veröffentlichten Forschungsdaten auf einer externen Plattform und nur etwa 2 % veröffentlichten Forschungsdaten als begleitende Dateien im institutionellen Repository der Leibniz Universität Hannover. Es wurde festgestellt, dass es statistisch signifikante Abhängigkeiten zwischen der Art der Forschungsdatenpublikation und der jeweiligen Fakultät, dem jeweiligen Zeitraum und der genutzten Sprache gibt. Es zeigte sich, dass der Anteil an extern publizierten Forschungsdaten über einen Zeitraum von zwölf Jahren, hauptsächlich in den letzten vier Jahren, stark angestiegen ist, sich jedoch zu großen Teilen auf nur drei Fakultäten beschränkt. Ähnlich zu diesem Wachstum in externen Forschungsdaten wechselte auch die dominante Sprache dieser Fakultäten von Deutsch auf Englisch. Da eine statistisch signifikante Abhängigkeit zwischen Sprache und Fakultät sowie Zeitgruppe nachgewiesen werden konnte, wurde die Sprache als eigenständiger Faktor unter Vorbehalt ausgeschlossen. Entsprechende mögliche Erklärungen zu den einzelnen signifikanten Relationen werden in Abschnitt 4.3 gegeben.

Insgesamt konnte festgestellt werden, dass nur ein sehr kleiner Anteil an Forschungsdaten so veröffentlicht wurde, dass sie den FAIR-Prinzipien entsprechen. Um diese Situation zu verbessern, wurden in Abschnitt 4.3 einige mögliche Handlungsempfehlungen für die LUH und die TIB ausgesprochen.

Abstract

This thesis deals with research data in dissertations and the associated research data management at German universities. The thesis is thematically divided into two parts.

In the first part, general and dissertation-specific guidelines of a sample of all German institutions entitled to confer doctoral degrees are examined for instructions on research data. This section is covered in Chapter 3.

In the second part, a representative, multi-layered sample of all dissertations from the institutional repository of the Leibniz University Hannover, published between 2012 and 2023, is examined with regard to research data. The focus is particularly on how this research data has been published: as an integrated part of the dissertation, as accompanying files, or as a separate publication on an external platform. This section is covered in Chapter 4.

The investigation of the general guidelines of the institutions entitled to confer doctoral degrees revealed that about 45 % of all institutions in Germany have a general research data policy (of these, about 60 % are universities). A further 45 % have at least a guideline on the rules of good scientific practice, while the remaining 10 % have no general policy mentioning research data or research data management (of these, about 71 % are art colleges). It was found that the type of institution and the type of sponsorship have a statistically significant influence on the highest classification level of the general documents. Corresponding possible explanations for the individual significant relationships are given in Section 3.3.1.

The investigation of the dissertation-specific guidelines of the institutions entitled to confer doctoral degrees revealed that only about 14 % of all institutions in Germany have at least one doctoral or examination regulation that obligatorily defines the handling of research data. Of these institutions, 75 % also have a general research data policy. More than half of all institutions have at least one doctoral or examination regulation containing mandatory regulations on the rules of good scientific practice, while about a quarter of all institutions have no doctoral or examination regulation containing research data or the rules of good scientific practice. It was found that the federal state, the type of institution, and the highest classification level of the general documents have a statistically significant influence on the highest classification level of the dissertation-specific documents. Corresponding possible explanations for the individual significant relationships are given in Section 3.3.2.

The investigation of dissertations from the institutional repository of the Leibniz University Hannover revealed that about 87 % of all dissertations in the sample are based at least in part on original primary data. Of these dissertations, about 61 % have published at least some of their research data. This proportion decreases to about 32 % when compressed photographs, spectral diagrams, gene sequences, unfilled questionnaires, guidelines, and assembly drawings are excluded due to lack of quality, relevance, or difficult-to-evaluate originality. The most common type of publication was research data integrated into the dissertation file, which accounted for about

57 % of all dissertations with original primary data. Only about 7 % of all dissertations with primary data published research data on an external platform, and only about 2 % published research data as accompanying files in the institutional repository of the Leibniz University Hannover. It was found that there are statistically significant dependencies between the type of research data publication and the respective faculty, the respective period, and the language used. It was shown that the proportion of externally published research data increased significantly over a period of twelve years, mainly in the last four years, but was largely confined to only three faculties. Similar to this growth in external research data, the dominant language in these faculties changed from German to English. Since a statistically significant dependency between language and faculty as well as time group could be demonstrated, language was excluded as an independent factor with reservation. Corresponding possible explanations for the individual significant relationships are given in Section 4.3.

Overall, it was found that only a very small proportion of research data was published in accordance with the FAIR principles. To improve this situation, several possible recommendations for action for LUH and TIB are given in Section 4.3.

Danksagung

Als erstes möchte ich meinen zwei Gutachtern, Dr. Sarah Dellmann und Prof. Dr. Robert Jäschke, für ihre Unterstützung sowie für ihre Bereitwilligkeit, dieses viel zu lange Dokument zu lesen und zu begutachten, danken.

Darüber hinaus möchte ich mich bei allen Mitarbeitenden der Technischen Informationsbibliothek bedanken, die sich bereit erklärt haben, mir in beratender Funktion zur Seite zu stehen und es mir auch ermöglicht haben, im Rahmen eines Praktikums administrativen Zugriff auf das institutionelle Repositorium der Leibniz Universität Hannover zu erhalten. Hier ist ein besonderer Dank an alle Mitglieder des Projekts *FoHop!* und an die Mitarbeitenden aus dem Bereich *Publikationsdienste und Open Access* gerichtet. Mein Dank an die Mitarbeitenden der Technischen Informationsbibliothek inkludiert dabei, alphabetisch nach Nachnamen sortiert, Anika Altenberg, Olga Engelhardt, Israel Holger, Petra Mensing, Linna Lu, Anna-Karina Renziehausen, Corinna Schneider, Cécilia Schröer und Anne Quinkenstein.

Auch bedanken möchte ich mich bei allen Mitarbeitenden der Universitätsbibliothek Johann Christian Senckenberg, die mich im Rahmen meines Bibliotheksreferendariats begleitet und betreut haben. Hier besonders zu erwähnen sind Bernhard Wirth, welcher es uns Referendaren durch seine sorgfältige und interaktive Planung des Referendariats erlaubte, alle Facetten des bibliothekarischen Alltags zu erleben, Agnes Brauer, welche mich stets offen und enthusiastisch in meinem Schwerpunkt, den *Digital Humanities*, betreut hat, Daniela Poth, welche als Direktorin der Bibliothek stets ein offenes Ohr für uns Referendare hatte und uns tiefgehende Einblicke in die Leitung einer wissenschaftlichen Bibliothek gewährte, und Dr. Christoph Marutschke, welcher mir meinen ersten tiefgehenden und praktischen Einblick in das Thema Forschungsdatenmanagement ermöglichte. Darüber hinaus gebührt mein Dank hier auch Michelle Kamolz, Jakob Frohmann und unzähligen weiteren Kolleginnen und Kollegen, die sich hier leider nicht alle aufzählen lassen.

Auch möchte ich mich bei meinen hessischen Mitreferendarinnen und Mitreferendaren für ihre Kollegialität, gegenseitige Unterstützung und Freundschaft bedanken. Dieses besondere Dank geht an Alicia Schwammborn, die es irgendwie geschafft hat, zwei Jahre mit mir an derselben Bibliothek auszuhalten, Maureen Bössow, Marvin Gusen und Henrike Weyer.

Mein Dank geht auch an das Institut für Bibliotheks- und Informationswissenschaft, an all seinen Mitarbeitenden, an alle meine Kommilitoninnen und Kommilitonen der Matrikel 28 sowie an alle Dozentinnen und Dozenten, die mir ein unglaublich abwechslungsreiches und spannendes Studium ermöglichten.

Auch möchte ich zutiefst bei meiner Frau, Sarah C. Krassnig, bedanken. Ich kann mir nur schwer eine Welt vorstellen, in der diese Abschlussarbeit in ihrer jetzigen Form hätte kreierte werden können, wenn sie mich nicht allzeit unterstützen würde.

Mein ewigwährender Dank geht auch an die beste Familie der Welt, die ich mir hätte wünschen können: Karl-Heinz, Roswitha, Stefan, Anna, Gerlinde, Anton und Therese Krassnig, Egon und Käthe Stein sowie Mary, Leyla und Melissa Sagin.

DIESE SEITE WURDE ABSICHTLICH LEER GELASSEN

Inhaltsverzeichnis

Zusammenfassung	v
Abstract	vii
Danksagung	ix
Tabellenverzeichnis	xiii
Abbildungsverzeichnis	xvii
Abkürzungsverzeichnis	xix
1 Einführung	1
2 Stand der Forschung	5
2.1 Grundlagen zu Forschungsdaten	5
2.1.1 Gute wissenschaftliche Praxis und FAIRe Forschungsdaten . .	6
2.1.2 Publikationsarten	7
2.1.3 Forschungsdatenrepositorien	9
2.1.4 Metadaten-Schemata	10
2.2 Richtlinien zu Forschungsdaten	10
2.3 Forschungsdaten in Dissertationen	11
3 Richtlinien zu Forschungsdaten aus deutschen Promotionsvorhaben	13
3.1 Material & Methoden	13
3.1.1 Material	13
3.1.2 Methoden	17
3.2 Resultate	18
3.2.1 Mögliche relevante Faktoren	18
3.2.2 Allgemeine Dokumente	19
3.2.3 Promotionspezifische Dokumente	22
3.3 Diskussion	25
3.3.1 Allgemeine Dokumente	26
3.3.2 Promotionspezifische Dokumente	28
4 Forschungsdaten im Repositorium der Leibniz Universität Hannover	31
4.1 Material & Methoden	31
4.1.1 Material	31
4.1.2 Methoden	34
4.2 Resultate	37
4.2.1 Nutzungsrate des LUH-Repos	38
4.2.2 Mögliche relevante Faktoren	39

4.2.3	Rate an erzeugten Primärdaten	39
4.2.4	Zeitliche Entwicklung der Forschungsdaten	40
4.2.5	Sprache und Forschungsdaten	41
4.2.6	Fakultäten und Forschungsdaten	42
4.2.7	Externe Publikationen und Metadaten	47
4.3	Diskussion	48
4.3.1	Sprache	48
4.3.2	Externe Forschungsdaten	49
4.3.3	Zeitliche Entwicklung von Forschungsdaten und Fakultäten .	49
4.3.4	Handlungsempfehlungen	50
5	Schlussfolgerungen	53
	Literaturverzeichnis	57
A	Richtlinien zu Forschungsdaten aus deutschen Promotionsvorhaben	69
A.1	Tabellen	69
B	Forschungsdaten im Repositorium der Leibniz Universität Hannover	71
B.1	Tabellen	71
B.2	Grafiken	73

Tabellenverzeichnis

3.1	Die Verteilung aller promotionsberechtigter Institutionen in Deutschland nach <i>Hochschultyp</i> \times <i>Trägerschaft</i> aufgegliedert. Absolute Werte in Klammern angegeben.	14
3.2	Die Verteilung der Institutionen in der Stichprobe nach <i>Hochschultyp</i> \times <i>Trägerschaft</i> aufgegliedert. Absolute Werte in Klammern angegeben.	15
3.3	Verteilung der höchsten Klassifizierungsstufe der Institutionen durch ihre allgemeingültigen Dokumente nach <i>Institutionstyp</i> \times <i>Klassifikationsstufe</i> aufgegliedert. Angabe relativ zum jeweiligen Institutionstyp. Absolute Werte in Klammern angegeben.	19
3.4	Verteilung der höchsten Klassifizierungsstufe der Institutionen durch ihre allgemeingültigen Dokumente nach <i>Trägerschaft</i> \times <i>Klassifikationsstufe</i> aufgegliedert. Angaben relativ zur jeweiligen Trägerschaft. Absolute Werte in Klammern angegeben.	19
3.5	Dokumentcharakter der FD-Richtlinien in relativer Angabe zum Dokumentnamen. Absolute Werte in Klammern angegeben.	22
3.6	Verteilung der höchsten Klassifizierungsstufe der Institutionen durch ihre promotionsspezifische Dokumente nach <i>Institutionstyp</i> \times <i>Richtlinienklassifikation</i> aufgegliedert. Angaben relativ zum jeweiligen Institutionstyp. Absolute Werte in Klammern angegeben.	23
3.7	Verteilung der höchsten Klassifizierungsstufe der Institutionen durch ihre promotionsspezifische Dokumente nach <i>Trägerschaft</i> \times <i>Richtlinienklassifikation</i> aufgegliedert. Angaben relativ zur jeweiligen Trägerschaft. Absolute Werte in Klammern angegeben.	23
3.8	Verteilung der höchsten Klassifizierungsstufe der Institutionen durch ihre promotionsspezifische Dokumente nach <i>Allgemeine Dokumentklassifikation</i> \times <i>Spezifische Dokumentklassifikation</i> aufgegliedert. Angaben relativ zum Gesamtwert der Stichprobeninstitutionen. Absolute Werte in Klammern angegeben.	24
3.9	Die Verteilung der Institutionen nach <i>Allgemeine Dokumentklassifikation</i> \times <i>Spezifische Dokumentklassifikation</i> aufgegliedert. Angaben relativ zu <i>Allgemeine Dokumentklassifikation</i> . Absolute Werte in Klammern angegeben.	25
3.10	Die Verteilung der Institutionen nach <i>Allgemeine Dokumentklassifikation</i> \times <i>Spezifische Dokumentklassifikation</i> aufgegliedert. Angaben relativ zu <i>Spezifische Dokumentklassifikation</i> . Absolute Werte in Klammern angegeben.	25
4.1	Die Verteilung der Grundmengen-Metadateneinträge nach <i>Fakultät</i> \times <i>Zeitraum</i> aufgegliedert. Absolute Werte in Klammern angegeben.	32

4.2	Die Verteilung der Stichproben-Metadateneinträge nach <i>Fakultät</i> \times <i>Zeitraum</i> aufgegliedert. Angaben relativ zur Gesamtsumme der Stichprobe. Absolute Werte in Klammern angegeben.	33
4.3	Die Verteilung der Stichproben-Metadateneinträge nach <i>Fakultät</i> \times <i>Zeitraum</i> aufgegliedert. Angaben relativ zu dem jeweiligen <i>Fakultät</i> \times <i>Zeitraum</i> -Wert der Grundmenge. Absolute Differenzwerte in Klammern angegeben.	34
4.4	Die Verteilung der Dissertationen laut den Zahlenspiegeln der LUH nach <i>Fakultät</i> \times <i>Zeitraum</i> aufgegliedert. Angaben relativ zu der Gesamtsumme der Zahlenspiegel-Dissertationen. Absolute Werte in Klammern angegeben. Spalten, die zumindest teilweise auf simulierten Werten basieren, sind mit einem Sternchen (*) markiert.	37
4.5	Der Anteil der Grundmenge nach <i>Fakultät</i> \times <i>Zeitraum</i> aufgegliedert relativ zu der respektiven <i>Fakultät</i> \times <i>Zeitgruppe</i> -Gesamtanzahl aller publizierten Dissertationen. Absolute Differenzwerte in Klammern angegeben. Spalten, die zumindest teilweise auf simulierten Werten basieren, sind mit einem Sternchen (*) markiert.	38
4.6	Anteil an PD-Dissertationen aus der Stichprobe. Angaben relativ zu der respektiven <i>Fakultät</i> \times <i>Zeitgruppe</i> -Gesamtanzahl. Absolute Werte in Klammern angegeben.	39
4.7	FD-Klassifizierung der PD-Dissertationen aus der Stichprobe nach <i>Publikationsart</i> \times <i>Klassifikationsstufe</i> \times <i>Jahresgruppe</i> aufgegliedert. Angaben relativ zu der Gesamtanzahl der jeweiligen Zeitgruppe. Absolute Werte in Klammern angegeben.	40
4.8	FD-Klassifizierung der PD-Dissertationen aus der Stichprobe nach <i>Sprache</i> \times <i>Zeitgruppe</i> aufgegliedert. Angaben relativ zu der Gesamtanzahl der jeweiligen Zeitgruppe. Absolute Werte in Klammern angegeben.	42
4.9	FD-Klassifizierung der PD-Dissertationen aus der Stichprobe nach <i>Fakultät</i> \times <i>Klassifikationsstufe</i> aufgegliedert. Angabe relativ zu der respektiven Gesamtanzahl für <i>Fakultät</i> . Absolute Werte in Klammern angegeben.	43
A.1	Klassifikation der allgemeingültigen Dokumente in relativer Angabe zum Bundesland. Absolute Werte in Klammern angegeben. Angabe der Bundesländer nach ISO 3166-2:2020 [51].	69
A.2	Klassifikation der promotionsspezifischen Dokumente in relativer Angabe zum Bundesland. Absolute Werte in Klammern angegeben. Angabe der Bundesländer nach ISO 3166-2:2020 [51].	70
B.1	FD-Klassifizierung der Dissertationen aus der Stichprobe nach <i>Sprache</i> \times <i>Zeitgruppe</i> aufgegliedert. Angaben relativ zu der Gesamtanzahl der jeweiligen Zeitgruppe. Absolute Werte in Klammern angegeben. .	71

B.2	FD-Klassifizierung der Dissertationen aus der Stichprobe nach <i>Fakultät</i> \times <i>Klassifikationsstufe</i> aufgegliedert. Angabe relativ zu der respektiven Gesamtanzahl für <i>Fakultät</i> . Absolute Werte in Klammern angegeben.	71
B.3	FD-Klassifizierung der Dissertationen aus der Stichprobe nach <i>Publikationsart</i> \times <i>Klassifikationsstufe</i> \times <i>Jahresgruppe</i> aufgegliedert. Angaben relativ zu der Gesamtanzahl der Jahresgruppe. Absolute Werte in Klammern angegeben.	72

DIESE SEITE WURDE ABSICHTLICH LEER GELASSEN

Abbildungsverzeichnis

2.1	Die Datenpublikationspyramide nach REILLY u. a. [3], basierend auf der Datenqualitätspyramide nach GRAY [28]. Eigene Übersetzung. . .	8
3.1	Die absolute Anzahl promotionsberechtigter Institutionen nach Bundesland. Steigende Farbintensivität korreliert mit steigender absoluten Anzahl.	14
3.2	Verteilung der Stichproben-Institutionen nach Bundesland. (A) Die absolute Anzahl der Institutionen nach Bundesland. (B) Der Anteil der Institutionen, die aus der Grundmenge übernommen wurden, relativ zum jeweiligen Bundesland. Steigende Farbintensivität zeigt relative Nähe zum Maximalwert an.	16
3.3	Verteilung der höchsten Klassifizierungsstufe der Institutionen durch ihre allgemeingültigen Dokumente nach <i>Bundesland</i> \times <i>Klassifikationsstufe</i> aufgegliedert. Angaben relativ zum jeweiligen Bundesland. Absolute Werte in Klammern angegeben. Angabe der Bundesländer nach ISO 3166-2:DE [51].	20
3.4	Anteil der verschiedenen Bezeichnungen für FD-Richtlinien. Absolute Werte in Klammern angegeben.	21
3.5	Verteilung der höchsten Klassifizierungsstufe der Institutionen durch ihre promotionsspezifische Dokumente in relativer Angabe zum Bundesland. Absolute Werte in Klammern angegeben. Angabe der Bundesländer nach ISO 3166-2:2020 [51].	24
4.1	Sprachen der PD-Dissertationen nach Fakultät und Zeitgruppe. Die Höhe der Balken entsprechen dem relativen Anteil zur jeweiligen angepassten <i>Fakultät</i> \times <i>Zeitgruppe</i> \times <i>Sprache</i> -Gesamtanzahl. Absolute Werte in Klammern angegeben.	43
4.2	FD für PD-Dissertationen nach Fakultät, Zeitgruppe und Klassifikationsstufe. Die Höhe der Balken entsprechen dem relativen Anteil zur jeweiligen angepassten <i>Fakultät</i> \times <i>Zeitgruppe</i> \times <i>FD</i> -Gesamtanzahl. Absolute Werte in Klammern angegeben.	44
4.3	Integrierte FD für PD-Dissertationen nach Fakultät, Zeitgruppe und Klassifikationsstufe. Die Höhe der Balken entsprechen dem relativen Anteil zur jeweiligen angepassten <i>Fakultät</i> \times <i>Zeitgruppe</i> \times <i>Integrierte FD</i> -Gesamtanzahl. Absolute Werte in Klammern angegeben.	45
4.4	Externe FD für PD-Dissertationen nach Fakultät, Zeitgruppe und Klassifikationsstufe. Die Höhe der Balken entsprechen dem relativen Anteil zur jeweiligen angepassten <i>Fakultät</i> \times <i>Zeitgruppe</i> \times <i>Begleitende FD</i> -Gesamtanzahl. Absolute Werte in Klammern angegeben.	46

4.5	Externe FD für PD-Dissertationen nach Fakultät, Zeitgruppe und Klassifikationsstufe. Die Höhe der Balken entsprechen dem relativen Anteil zur jeweiligen angepassten <i>Fakultät</i> \times <i>Zeitgruppe</i> \times <i>Externe FD</i> -Gesamtanzahl. Absolute Werte in Klammern angegeben.	47
B.1	FD-Klassifikation der Dissertationen aus der Stichprobe nach Fakultät, Zeitgruppe und Klassifikationsstufe. Die Höhe der Barren entsprechen dem relativen Anteil zur jeweiligen <i>Fakultät</i> \times <i>Zeitgruppe</i> -Gesamtanzahl. Absolute Werte in Klammern angegeben.	73

Abkürzungsverzeichnis

- BMBF** Bundesministerium für Bildung und Forschung 6, 13, 27
- DFG** Deutsche Forschungsgemeinschaft 5, 6, 27
- DMP** Datenmanagementplan 11, 52
- DNB** Deutsche Nationalbibliothek 12
- DOI** Digital Object Identifier 10, 35, 48
- FAIR** Findable, Accessible, Interoperable und Reusable 1, 6–9, 12, 21, 51, 55
- Fakultät ARC** Fakultät für Architektur und Landschaft 32–34, 37–39, 43–47, 49, 50, 54, 71, 73
- Fakultät BAU** Fakultät für Bauingenieurwesen und Geodäsie 32–34, 37–40, 43–47, 49, 50, 54, 71, 73
- Fakultät INF** Fakultät für Elektrotechnik und Informatik 32–34, 37–40, 43–47, 49, 54, 71, 73
- Fakultät JUR** Juristische Fakultät 32–34, 37–39, 43, 50, 54, 71, 73
- Fakultät MAS** Fakultät für Maschinenbau 32–34, 37–40, 43–47, 49, 50, 54, 71, 73
- Fakultät MAT** Fakultät für Mathematik und Physik 32–34, 37–40, 43–47, 54, 71, 73
- Fakultät NAT** Naturwissenschaftliche Fakultät 32–34, 37–40, 43–47, 49, 50, 54, 71, 73
- Fakultät PHI** Philosophische Fakultät 32–34, 37–39, 43–47, 49, 50, 54, 71, 73
- Fakultät WIWI** Wirtschaftswissenschaftliche Fakultät 32–34, 37–39, 43–47, 50, 54, 71, 73
- FD** Forschungsdaten xiii–xv, xvii, xviii, 1–12, 16, 17, 19–29, 31, 33–37, 39–55, 69–73
- FDM** Forschungsdatenmanagement 1–6, 11–13, 15–17, 21, 25, 26, 28, 29, 48, 50–52, 54
- FH** Fachhochschule 13–15, 19, 23, 26
- GWK** Gemeinsame Wissenschaftskonferenz 5, 16
- GWP** Gute wissenschaftliche Praxis 6, 11, 16, 17, 19–26, 53, 54, 69, 70

- HAW** Hochschule für Angewandte Wissenschaften 13–15, 19, 23, 26
- HRK** Hochschulrektorenkonferenz 5, 13
- HSeT** Hochschule eigenen Typs 14, 15, 19, 23
- KH** Künstlerische Hochschule 13–15, 19, 23, 26
- KI** Künstliche Intelligenz 51
- LUH** Leibniz Universität Hannover xiv, 3, 4, 31, 32, 34, 36–38, 50–52
- LUH-Repo** Institutionelles Repositorium der Leibniz Universität Hannover 3, 4, 31, 33, 34, 36–39, 47–51, 53–55
- NFDI** Nationale Forschungsdateninfrastruktur 5, 6
- OAI-PMH** Open Archives Initiative Protocol for Metadata Harvesting 31
- PD** Primärdatum 31, 34, 35, 37, 39, 40, 51
- PD-Dissertation** Dissertation mit produzierten Primärdaten (mit oder ohne Publikation dieser Primärdaten) xiv, xvii, xviii, 39–49, 51, 54, 55
- RfII** Rat für Informationsinfrastrukturen 5, 6
- TIB** Technische Informationsbibliothek / Leibniz-Informationszentrum Technik und Naturwissenschaften und Universitätsbibliothek 3, 4, 50, 51
- VH** Verwaltungshochschule 14, 15

Kapitel 1

Einführung

Die Veröffentlichung, langfristige Verfügbarkeit sowie allgemeine Nutzbarkeit von Forschungsdaten (FD) nach den FAIR-Prinzipien [1] gewinnen in der Wissenschaft zunehmend an Bedeutung [2]. Dies wird, zumindest in Teilen, durch ein wachsendes Bewusstsein für die Bedeutung von Transparenz und Nachvollziehbarkeit wissenschaftlicher Ergebnisse und den damit einhergehend zunehmenden Forderungen nach Open Access und Open Data Initiativen getrieben. So rückt das Thema Forschungsdatenmanagement (FDM) zunehmend in den Fokus der Wissenschaft.

Einleitung

Entsprechend dieser stetig wachsenden Bedeutung von FDM sind in der Bibliotheks- und Informationswissenschaft folgende Themen vermehrt Gegenstand präskriptiver und deskriptiver Forschung: (i) wie FD langfristig gelagert und zugänglich gemacht werden können und wie diese bisher tatsächlich gelagert und zugänglich gemacht werden, (ii) welche Metadatenklassifikationssysteme hierzu genutzt werden sollten und welche in der Praxis bisher genutzt werden, (iii) welche Dateitypen für langfristige Lagerung und Zugänglichkeit geeignet sind und welche nicht, (iv) welche Möglichkeiten es gibt, den richtigen Umgang mit FD durch die Forschenden zu fördern und (v) wie eine entsprechende Infrastruktur aufgebaut werden kann und welche Infrastruktur bereits existiert.

*Bisherige
Forschung*

Da die Klassifikation von FD für den Aufbau und die Bearbeitung dieser Abschlussarbeit von sehr hoher Relevanz ist, ist hier spezifisch zu erwähnen, dass sich in der Literatur eine hierarchische Klassifikationsmöglichkeit für FD nach Publikationsart herauskristallisiert hat. Diese evaluiert, inwiefern die FD mit dem dazugehörigen Text verknüpft bzw. in diesen integriert worden sind. So lässt sich zwischen drei verschiedenen Publikationsformen für FD unterscheiden [3, S. 36ff.]:

Klassifikation

1. vollständig integrierte Daten (z.B. Tabellen und Grafiken, die in der PDF-Datei eingebettet worden sind),
2. beigefügte Daten (z.B. Dateien, die gemeinsam mit der PDF-Datei der Publikation auf dem Publikationsserver hochgeladen worden sind)
3. Daten, die separat zu der dazugehörigen Publikation auf einer Plattform für FD hochgeladen werden (z.B. fachspezifische FD-Repositoryen)

Für Details zu dieser Klassifikationshierarchie, siehe Kapitel 2.

Die meiste Literatur zum Thema FDM ist eher allgemeiner Natur oder hat ein Hauptaugenmerk auf jene FD, welche mit Artikeln in wissenschaftlichen Journals produziert wurden [4]. Ein zumeist wenig beachteter Aspekt sind jene Daten, die im Rahmen des Verfassens einer Hochschulschrift generiert worden sind und daher nicht primär mit einem Journalartikel assoziiert werden. Dies bedeutet allerdings

Wissenslücke

nicht, dass zu diesem Thema eine vollständige Forschungsarmut besteht. Jedoch sind Publikationen zu diesem Thema, zumindest bisher, meist eher präskriptiver statt deskriptiver Natur: So geben im wissenschaftlichen Kontext z.B. präskriptive Publikationen aus dem DFG-Förderprojekt „eDissPlus“ [5, 6, 7] sowie die „Policy für dissertationsbezogene FD“ der Deutschen Nationalbibliothek [8] vermehrt Richtlinien für den Umgang mit FD in Hochschulschriften. Es fehlen bisher allerdings umfassende Studien zur Wirksamkeit bzw. Durchsetzung dieser Richtlinien bei Studierenden und Nachwuchsforschenden (z.B. durch entsprechende Prüfungsordnungen und Beratungsangebote zu diesem Thema durch Universitätsbibliotheken). Hier existieren bisher größtenteils nur spezialisierte bzw. fachspezifische Untersuchungen [9] oder Studien, die nur integrierte und begleitende FD aus dem Appendix der jeweiligen Dissertationen untersuchen [10].

Motivation

Dabei sind Hochschulschriften–insbesondere Dissertationen–und deren Erschließung, Publikation und Archivierung ein zentraler Bestandteil jener Dienstleistungen, die an Hochschulbibliotheken anfallen. So sind Dissertationen, neben Habilitationsschriften, die wichtigsten Hochschulschriften, die an Forschungsinstitutionen produziert werden. Noch dazu werden diese Schriften zunehmend in institutionellen Repositorien veröffentlicht, welche wiederum zumeist entweder von der wissenschaftlichen Bibliothek betrieben oder zumindest mitbetreut werden. Entsprechend ist eine fachgerechte, langfristige sowie zugängliche FD-Infrastruktur, die sich an relevante internationale Standards hält, sowie passende Beratungsgebote für Forschende zum Thema FDM in Dissertationen unabdingbar. Um dies effizient und zufriedenstellend bewerkstelligen zu können, benötigen Bibliothekare jedoch ausreichende empirisch-deskriptive Daten sowie präskriptive Richtlinien, an denen sie sich orientieren und von denen sie ihr lokales Beratungs- und Infrastrukturangebot ableiten können [11]. Daher ist es unabdingbar, diese Lücke in der wissenschaftlichen Literatur sukzessive zu minimieren, um die Effizienz und Relevanz bibliothekarischer Arbeit in diesem Bereich zu optimieren.

Forschungsfragen

Um etwas zu diesen Bemühungen beizusteuern, stellt diese Arbeit zwei thematisch zentrale Forschungsfragen, welche sich respektiv in zwei oder fünf untergeordnete Forschungsfragen untergliedern lassen. Hierbei konzentriert sich diese Arbeit bei den Hochschulschriften exklusiv auf *Dissertationen*, da diese durch ihre wissenschaftlich hohe Relevanz und Frequenz, den wichtigsten Teil der Hochschulschriften ausmachen. Die Forschungsfragen lauten wie folgt:

1. Inwiefern wird der Umgang mit FD für Dissertationen in Deutschland bereits in verbindlichen verwaltungsrechtlichen Dokumenten geregelt?
 - a) Inwiefern wird der Umgang mit FD in den allgemeinen Richtlinien einer Institution geregelt, welche auch Promovierende und ihr Forschungsvorhaben betreffen?
 - b) Inwiefern wird der Umgang mit FD in Promotions- und Prüfungsordnungen geregelt?

-
2. Auf welche Art und Weise haben Promovierende, welche ihre Dissertationen 2012–2023 in einem institutionellen Repository publiziert haben, ihre FD in ihre Publikation integriert?
- Für welchen Anteil an Dissertationen wurden FD in welcher Form nach dem Klassifikationssystem von REILLY u. a. [3] veröffentlicht?
 - Inwiefern hat sich das Publikationsverhalten zu FD in Dissertationen in den letzten zwölf Jahren verändert?
 - Inwiefern unterscheiden sich die verschiedenen wissenschaftlichen Fachrichtungen in ihrem Publikationsverhalten in Bezug auf FD aus Dissertationen?
 - Wie wird in den Metadaten von FD aus Dissertationen sichtbar gemacht, dass es eine dazugehörige Dissertation gibt?
 - Wie wird in den Metadaten von Dissertationen sichtbar gemacht, dass es dazugehörige FD gibt?

Die erste Forschungsfrage behandelt die deutschlandweite Situation zu FD in Dissertationen aus verwaltungsrechtlicher Sicht, während die zweite zentrale Forschungsfrage dieser Arbeit das tatsächliche Publikationsverhalten von Promovierenden in Bezug auf FD in den letzten zwölf Jahren untersucht.

Der Zeitraum 2012–2023 wurde ausgesucht, da 2012 das FD-Repositorysverzeichnis *re3data* gegründet wurde [12] und die darauffolgenden Jahre von vermehrtem Interesse am korrekten Umgang mit FD geprägt waren (siehe Kapitel 2 für Details). Da eine umfassende und differenzierte Auswertung aller in diesem Zeitraum in Deutschland erschienenen Dissertationen im Rahmen einer Masterarbeit zur Beantwortung der zweiten Forschungsfrage nicht machbar wäre, konzentriert sich diese Arbeit explizit nur auf jene Dissertationen, die an der Leibniz Universität Hannover (LUH) im institutionellen Repository der Leibniz Universität Hannover (LUH-Repo) der Technischen Informationsbibliothek (TIB) veröffentlicht wurden. Dieses Repository wurde wegen zwei Gründen ausgewählt: Der erste Grund war, dass sich die TIB zum Verfassungszeitpunkt dieser Abschlussarbeit im Rahmen eines Projektes *FoHop!*–damit beschäftigt hat, inwiefern sich FD bereits im LUH-Repo befinden und wie diese gekennzeichnet werden können. Dadurch war es im Rahmen dieser Abschlussarbeit möglich, administrativen Zugriff auf das LUH-Repo zu erhalten und direkt einen praktischen Zweck für die Ergebnisse dieser Arbeit zu finden. Der zweite Grund war, dass das LUH-Repo aufgrund seiner großen Anzahl an vorhandenen Dissertationen sowie wegen seines breiten fachlichen Spektrums, welches von den großen Fächern hauptsächlich nur die Medizin und darstellende Kunst vermisst, ein fast ideales repräsentatives institutionelles Repository darstellt.

Diese Masterarbeit verfolgt einen mehrstufigen methodischen Ansatz zur Beantwortung der einzelnen Forschungsfragen. Die Arbeit ist hierfür in mehrere Kapitel gegliedert, die jeweils spezifische Aspekte der Forschungsfragen untersuchen und beantworten. In Kapitel 2 wird zunächst der aktuelle Stand der Forschung zu FD und FDM, insbesondere in Bezug auf Dissertationen, dargestellt. In Kapitel 3 werden

*Gründe der
Auswahl*

*Ziele und
Dokument-
struktur*

verwaltungsrechtliche Dokumente einer repräsentativen Stichprobe aller promotionsberechtigter Institutionen aus Deutschland analysiert. Der Fokus liegt dabei auf Richtlinien und Anforderungen, die den Umgang mit FD betreffen. Dies umfasst allgemeine Dokumente (z.B. FD-Richtlinien) und promotionspezifische Dokumente (z.B. Promotions- und Prüfungsordnungen). In Kapitel 4 wird das LUH-Repo und die sich darin befindenden Dissertationen in Bezug auf primäre FD, die während des dazugehörigen Forschungsprozesses generiert worden sind, untersucht. Es wird geprüft, wie die Fakultäten der LUH sich im Umgang mit FD unterscheiden und wie sich deren Praxis über einen Zeitraum von zwölf Jahren entwickelt hat. Auch werden hier spezifische Handlungsempfehlungen für die LUH und die TIB in Bezug auf FD und FDM formuliert. Abschließend werden in Kapitel 5 die Erkenntnisse aus den vorherigen Kapiteln zusammengeführt und die Forschungsdaten noch einmal explizit beantwortet. Der Aufbau der einzelnen Kapitel wird jeweils zu Beginn des jeweiligen Kapitels detailliert.

Kapitel 2

Stand der Forschung

In diesem Kapitel wird der aktuelle Stand der Forschung für diese Abschlussarbeit zusammengefasst. Hiermit kann der Beitrag dieser Abschlussarbeit in dem breiteren wissenschaftlichen Kontext eingeordnet werden.

Thema des Kapitels

Hierzu werden in Abschnitt 2.1 allgemeine Informationen zu FD und FDM dargestellt. Dies umfasst, wie FD definiert und klassifiziert werden, welche Prinzipien und Standards nach heutiger Auffassung in Relation zu FD befolgt werden sollten und wie FD aus empirischer Sicht bisher veröffentlicht wurden. Darauf folgend werden in Abschnitt 2.2 bisherige Studien zu dem Thema FD-Richtlinien in Deutschland zusammengefasst. Schließlich wird in Abschnitt 2.3 eine Übersicht gegeben, inwiefern das Publizieren von FD im Rahmen von Dissertationen bereits erforscht wurde und was die dazugehörigen Ergebnisse waren.

Aufbau des Kapitels

2.1 Grundlagen zu Forschungsdaten

FD lassen sich wie folgt definieren:

Definition

„Unter digitalen Forschungsdaten verstehen wir [...] alle digital vorliegenden Daten, die während des Forschungsprozesses entstehen oder ihr Ergebnis sind.“ [13, S. 130]

Sie umfassen also alle (digitalen) Daten, die als Teil des wissenschaftlichen Prozesses entstanden sind. Dies inkludiert Mess-, Befragungs- und Beobachtungsergebnisse, umfasst aber auch methodische Komponenten, die für die Reproduzierbarkeit der Ergebnisse vonnöten sind, wie z.B. (unbeantwortete) Fragebögen und Leitfäden [14].

Seit etwa zehn Jahren rückt der korrekte und nachhaltige Umgang mit FD zunehmend in den Fokus der deutschen Wissenschaft. Dies lässt sich z.B. durch die wissenschaftspolitischen Entwicklungen der letzten Jahre erkennen.

Wichtigkeit

So betonte die Hochschulrektorenkonferenz (HRK) bereits 2014, dass das exponentielle Wachstum und die Komplexität digitaler Forschungsdaten neue Methoden und Werkzeuge erfordern. Hier forderte die HRK Hochschulleitungen explizit dazu auf, Leitlinien zu entwickeln, Informationskompetenz zu fördern und effizientes FDM zu unterstützen. Dies solle vom Bund und den Ländern durch übergreifende Maßnahmen koordiniert und zumindest in Teilen finanziert werden [15]. Nur wenig später hat auch der, von der Gemeinsamen Wissenschaftskonferenz (GWK) beschlossene, Rat für Informationsinfrastrukturen (RfII) entsprechende Empfehlungen veröffentlicht und vorgeschlagen, dass eine Nationale Forschungsdateninfrastruktur (NFDI) aufgebaut werden sollte [16]. Dies wurde von der Deutschen Forschungsgemeinschaft (DFG)

Nationale Forschungsdateninfrastruktur

unterstützt [17] und endete 2018 in einer Vereinbarung zwischen Bund und Ländern, eine entsprechende Infrastruktur aufzubauen [18]. Hieraus resultierte die Gründung des NFDI e.V. in 2020 [19].

FD-Gesetz Darüber hinaus wird zum Zeitpunkt dieser Abschlussarbeit vom Bundesministerium für Bildung und Forschung (BMBF) angestrebt, dass ein FD-Gesetz verabschiedet wird, welches den Zugang zu Daten aus öffentlicher Hand erleichtern, Datenschutzregelung zu Gunsten besserer FD-Handhabung verändern, FD durch neue Metadatenkataloge der Forschungseinrichtungen auffindbarer machen und ein *Micro Data Center* für Statistik- und Registerdaten aufbauen soll [20]. Dieses Vorhaben wird auch sowohl von der DFG [21] wie auch von dem RfII [22] unterstützt.

Meinung der Wissenschaftler Entsprechend ändert sich auch langsam die Bereitwilligkeit von Forschenden, ihre FD anderen Personen zugänglich zu machen [23]. So zeigte z.B. eine Studie in 2019, dass nur 11 % älterer Forschenden im Bereich der Biologie gewillt sind, Daten auf Anfrage anderer Forschenden zur Verfügung zu stellen, während dieselbe Rate der Nachwuchsforschenden aus derselben Disziplin bei 72 % liegt [24].

Aufbau des Abschnitts Der korrekten wissenschaftlichen Handhabung von FD und FDM wird also sehr hohe Bedeutung beigemessen. In diesem Abschnitt werden hierzu die grundlegenden Informationen zum aktuellen Stand wiedergegeben. In Abschnitt 2.1.1 werden die Regeln der guten wissenschaftlichen Praxis (GWP) und die FAIR-Prinzipien, auf die in der GWP referenziert werden, erklärt. In Abschnitt 2.1.2 werden die unterschiedlichen Publikationsarten für FD dargelegt. Die Nutzung von externen FD-Repositoryn wird in Abschnitt 2.1.3 eingehender erklärt. In Abschnitt 2.1.4 werden die am häufigsten genutzten Metadaten-Schemata, die für FD verwendet werden, grob umrissen.

2.1.1 Gute wissenschaftliche Praxis und FAIRe Forschungsdaten

GWP-Regeln Als Grundlage für das wissenschaftliche Verhalten in Deutschland gelten die Regeln der GWP der DFG [25]. Diese spezifizieren korrektes wissenschaftliches Verhalten für Forschungsvorhaben und wie deren Umgebung gestaltet werden sollte. Dies reicht von organisatorischen Angelegenheiten (z.B. die Leitung wissenschaftlicher Einrichtungen) bis hin zu operativen Anforderungen (z.B. die phasenübergreifende Qualitätssicherung innerhalb eines Forschungsprojektes). Da FD ein elementarer Bestandteil des wissenschaftlichen Prozesses sind, geben die Regeln der GWP entsprechende Leitlinien für deren Produktion, Bearbeitung und Publikation vor. So sollten z.B., gemäß *Leitlinie 7* der GWP, die Generierung, Prozessierung und Analyse von FD unter Einhaltung fachspezifischer Standards und Methoden vollzogen und deren Umfang und Art ausreichend dokumentiert werden. Von besonderem Interesse für diese Abschlussarbeit ist dabei *Leitlinie 13* der GWP, welche besagt, dass öffentlicher Zugang zu den Forschungsergebnissen, gemäß den sogenannten FAIR-Prinzipien, hergestellt werden sollte.

FAIRe Daten Die FAIR-Prinzipien besagen, dass FD *Findable* (dt. *Auffindbar*), *Accessible* (dt. *Zugänglich*), *Interoperable* (dt. *Interoperabel*) und *Reusable* (dt. *Wiederverwendbar*) sein sollten [1]. Diese Richtlinien lassen sich wiederum in einzelne Unterempfehlungen

aufgliedern, wie das entsprechende Ziel erreicht werden sollte oder was notwendig ist, um diese Ziel erreichen zu können. Diese Unterempfehlungen werden durch den jeweiligen Anfangsbuchstaben des FAIR-Akronyms und einer aufsteigenden Ziffer voneinander differenziert.

Unter *Findable* wird Forschern empfohlen, dass Daten und dazugehörige Metadaten so beschrieben werden sollten, dass sie sowohl für Menschen als auch für Maschinen leicht auffindbar sind. Um dies zu erreichen sollten Daten und Metadaten eine global eindeutige und dauerhafte Kennung erhalten (F1), Daten mit umfangreichen Metadaten beschrieben werden (F2), Metadaten eindeutig und explizit die Kennung der beschriebenen Daten enthalten (F3) und (Meta-)Daten in einer durchsuchbaren Ressource registriert oder indiziert worden sein (F4).

Findable

Unter *Accessible* wird verstanden, dass, die Daten, einmal auffindbar, auch leicht zugänglich sein sollten. Dies bedeutet, dass die Daten so zugänglich gemacht werden, dass diese unter Angabe der (Meta-)Datenkennung durch ein standardisiertes Kommunikationsprotokoll abgerufen werden können (A1) und auf die Metadaten auch dann zugegriffen werden kann, wenn die Daten nicht mehr verfügbar sein sollten (A2). Hierbei sollte das Protokoll offen sowie universell implementierbar sein (A1.1) und, bei Bedarf, ein Authentifizierungs-/Autorisierungsverfahren ermöglichen (A1.2).

Accessible

Unter *Interoperable* versteht man, dass (Meta-)daten in Formaten vorliegen sollten, die die Integration und das Zusammenspiel mit anderen Daten und Anwendungen ermöglicht. Es soll eine zugängliche, gemeinsame, formale und allgemein anwendbare Sprache für die Wissensrepräsentation in den (Meta-)Daten genutzt werden (I1), Vokabulare verwendet werden, die den FAIR-Prinzipien folgen (I2) und die (Meta-)Daten sollen qualifizierte Verweise auf andere (Meta-)Daten enthalten (I3).

Interoperable

Schlussendlich besagt *Reusable*, dass die Daten so aufbereitet und dokumentiert sein sollten, dass sie von anderen Forschern unter den angegebenen Bedingungen wieder- und weiterverwendet werden können. Dies erfordert detaillierte Beschreibungen der Datenquelle, des Datenerstellungsprozesses und der zugrunde liegenden Methoden (R1.2), einer eindeutigen und zugänglichen Nutzungslizenz (R1.1) und die Nutzung domänenrelevanter Community-Standards (R1.3).

Reusable

Hierbei sollte jedoch erwähnt werden, dass freier Zugriff auf Forschung und deren Ergebnisse unter ethischen und rechtlichen Einschränkungen zwar prinzipiell empfehlenswert ist [26], und auch einen gewissen Zitationsvorteil für die Publizierenden bietet [4, 27], jedoch nicht Teil der FAIR-Prinzipien ist.

FAIR und Open Data

2.1.2 Publikationsarten

In der Literatur zu FD wird im Allgemeinen, basierend auf der Integrationsstufe der FD zu einem dazugehörigen wissenschaftlichen Schriftwerk, zwischen drei verschiedenen Publikationsarten unterschieden [3]. Diese Stufen werden mit der sogenannten *Publikationspyramide* dargestellt, siehe Abbildung 2.1, welche von oben nach unten die drei Publikationsarten mit abnehmender Integrationsstärke zeigt. Wobei die unterste Stufe unveröffentlichte Datensätze repräsentiert [3].

Publikationspyramide

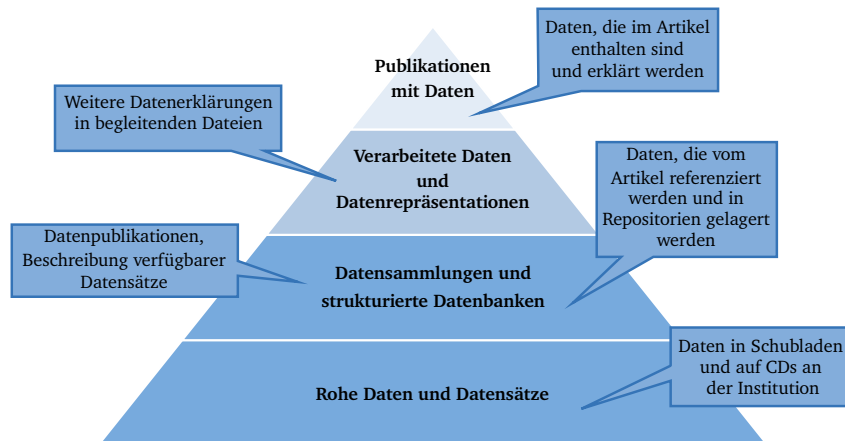


Abbildung 2.1: Die Datenpublikationspyramide nach REILLY u. a. [3], basierend auf der Datenqualitätspyramide nach GRAY [28]. Eigene Übersetzung.

Integrierte FD Entsprechend der Datenpublikationspyramide, sind die integriertesten FD jene, die als Teil eines wissenschaftlichen Schriftstückes veröffentlicht werden, ohne dass eine Trennung zwischen Dokument und FD stattfindet (z.B. Tabellen und Grafiken in einer PDF-Datei). Sie besitzen keine eigene Kennung und sind nicht über eigene Metadaten weiter erschlossen [3]. Entsprechend sind sie nach den FAIR-Prinzipien im Allgemeinen nicht *findable*, *accessible*, *interoperable* oder *reusable* [vgl. 3]. In dieser Abschlussarbeit werden solche FD als *integrierte FD* bezeichnet.

Begleitende FD Die nächste Publikationsart ist, wenn FD als begleitende Dateien eines wissenschaftlichen Schriftstückes auf derselben Plattform publiziert werden. Hierbei besitzen die FD typischerweise keine eigene Kennung oder erschließende Metadaten. Sie können diese jedoch in wenigen Fällen vorweisen [3]. Während die Situation hier bereits besser ist als bei den integrierten Daten, entsprechen diese Daten meist auch unzureichend den FAIR-Prinzipien [vgl. 3]. In dieser Abschlussarbeit werden solche FD als *begleitende FD* bezeichnet.

Externe FD Die dritte und letzte Publikationsart ist, wenn FD separat zu dem dazugehörigen Schriftstück auf einer anderen Plattform–oder zumindest unter einer anderen Kennung auf derselben Plattform–publiziert werden. In diesen Fällen werden die Daten im wissenschaftlichen Schriftstück als externer Datensatz zitiert und befinden sich idealerweise in dedizierten FD-Repositorien (siehe Abschnitt 2.1.3). Dies erlaubt eine eingehendere Beschreibung und Indexierung der Daten über entsprechende Metadaten (siehe Abschnitt 2.1.4) Auch möglich, jedoch bedeutend seltener, ist der Fall, dass die FD in entsprechenden FD-Journalen publiziert werden, die eine ähnliche Erschließung und Beschreibung erlauben. In solchen Fällen verweisen die Metadaten des Datensatzes wie auch die des Schriftstückes auf die Existenz des entsprechenden Konterparts [3]. Sie entsprechen, zu unterschiedlich starken Graden, den FAIR-Prinzipien [vgl. 3]. In dieser Abschlussarbeit werden solche FD als *externe FD* bezeichnet.

Die letzte Stufe der Pyramide betrifft dann jene FD, die nicht publiziert worden sind und damit auch keine (offizielle) Beziehung zu einem wissenschaftlichen Schriftstück besitzen. Diese Daten befinden sich meistens auf lokalen Servern oder anderen Datenträgern einer wissenschaftlichen Institution [3].

*Nicht
publiziert*

2.1.3 Forschungsdatenrepositorien

Um eine FAIRe Publizierung von FD zu erlauben, wurden über die Jahre einige Repositorien gegründet, deren dedizierter Zweck es ist, FD langfristig zu lagern, verfügbar und auffindbar zu machen. Während diese Repositorien früher stark verstreut waren und es schwer war, einen einigermaßen vollständigen Überblick über diese zu erhalten, wurde vor zwölf Jahren ein Großteil dieser Repositorien mit dem offenen *re3data*-Register in einem zentralen Index auffindbar und nach eigenen Bedürfnissen filterbar gemacht [12].

re3data

FD-Repositorien lassen sich in drei Kategorien einteilen: fachspezifische Repositorien, institutionelle Repositorien und allgemeine Repositorien.

Kategorien

Unter fachspezifischen Repositorien werden jene Repositorien verstanden, auf denen nur FD einer bestimmten wissenschaftlichen Disziplin hochgeladen werden dürfen. Diese machen mit ungefähr zwei Drittel ($n = 2212$) aller Repositorien auf *re3data* ($n = 3274$) den Großteil aller Repositorien aus [29].

Fachspezifisch

Unter institutionellen Repositorien werden jene Repositorien verstanden, auf denen nur FD hochgeladen werden, welche an der dazugehörigen Forschungsinstitution produziert wurden oder in Kooperation mit besagter Institution entstanden sind. Diese sind hierbei jedoch meist disziplinübergreifend. Institutionelle Repositorien machen ungefähr ein Fünftel ($n = 761$) aller Repositorien auf *re3data* ($n = 3274$) aus [29]. Von diesen sind wiederum 149 institutionelle Repositorien aus Deutschland [30], womit Deutschland, nach den Vereinigten Staaten von Amerika, die zweithöchste Anzahl an institutionellen Repositorien aller Länder hat.

Institutionell

Unter allgemeinen Repositorien werden jene Repositorien verstanden, welche disziplinübergreifend FD aufnehmen, jedoch nicht an eine bestimmte Forschungsinstitution gebunden sind. Die bekanntesten Beispiele hierfür sind *Zenodo*, *Dryad* und *Figshare*.

Allgemein

Als Basis für FD-Repositorien dienen sowohl speziell dafür entwickelte Software wie auch allgemeine Datenbanksysteme. So ist die am häufigsten genutzte Software für FD-Repositorien *Dataverse*, dicht gefolgt von *DSpace* [29]. Diese machen respektiv 11,11 % ($n = 119$) und 10,27 % ($n = 110$) aller Repositorien auf *re3data*, die Informationen zu der genutzten Software besitzen ($n = 1071$), aus (eigene Berechnung auf Basis von Daten aus [29]). Das am häufigsten genutzte allgemeine Datenbanksystem–und das dritthäufigst genutzte System insgesamt–ist *MySQL* mit 8,03 % ($n = 86$; eigene Berechnung auf Basis von Daten aus [29]).

Software

Allen FD-Repositorien gemein ist, dass sie durch Metadaten erlauben, die hochgeladenen FD weiter zu beschreiben und zu erschließen.

Metadaten

2.1.4 Metadaten-Schemata

Standards Für die Erschließung von Metadaten gibt es einige Standards, welche es über multiple Repositorien hinweg erlauben sollten, Daten uniform beschreiben und suchen zu können. Die am häufigsten genutzten Metadatenschemata für diesen Zweck sind hierbei *Dublin Core* [31] und *DataCite Metadata Schema* [32], welche respektiv 44,37 % ($n = 595$) und 32,44 % ($n = 435$) aller *re3data*-Repositorien mit Einträgen zu Metadatenschemata mit einem Eigennamen ($n = 1341$) ausmachen [29, 33]. Eine simultane Nutzung mehrerer Metadatenschemata ist möglich.

Dublin Core *Dublin Core* ist ein Metadatenschema, das entwickelt wurde, um die Beschreibung von Ressourcen im Internet zu vereinheitlichen und zu vereinfachen. Das Schema ist hierbei so konzipiert, dass es nur aus grundlegenden Elementen besteht, die vielseitig einsetzbar sind. Es besteht aus einem Satz von 15 grundlegenden Elementen (z.B. Titel, Autor, Thema etc.), die auf eine Vielzahl von Ressourcen angewandt werden können. Im *Dublin Core* Metadata Standard werden Einträge zwischen FD und den dazugehörigen Schriftstücken jeweils durch *dcterms:isPartOf* und *dcterms:hasPart* verzeichnet.

DataCite Das *DataCite Metadata Schema*, hingegen, wurde explizit für FD entwickelt und beinhaltet eine differenziertere Anzahl an möglichen Metadatenelementen ($n = 22$) mit einem kontrollierten Vokabular für viele seiner Einträge. Im *DataCite Metadata Schema* werden Einträge zwischen FD und den dazugehörigen Schriftstücken typischerweise jeweils durch ein *relatedIdentifier*-Eintrag mit dem *relationType IsSupplementTo* und *IsSupplementedBy* verzeichnet [34, 35]–idealerweise unter Angabe der entsprechenden Digital Object Identifiers (DOIs). Allerdings wird in einigen Fällen auch der *relationType IsReferencedBy* bzw. *References* genutzt, insofern dies auf die FD und das dazugehörige Schriftstück zutrifft [35].

Konversion Eine Zuordnung zwischen den beiden Metadatenschemata wurde von der *DataCite Metadata Working Group* erstellt [36]. Hiermit ist eine Konversion zwischen den beiden Standards möglich–allerdings besteht die Möglichkeit von Informationsverlust, wenn von *DataCite Metadata Schema* zu *Dublin Core* konvertiert wird, da letzteres nur weniger differenziert Informationen encodieren kann [36].

2.2 Richtlinien zu Forschungsdaten

Wachstumstrend Entsprechend der wachsenden Bedeutung vom korrekten Umgang mit FD haben über die letzten zehn Jahre vermehrt Institutionen FD-Richtlinien verabschiedet, die den korrekten Umgang mit FD regeln oder Empfehlungen hierzu aussprechen. So gab es 2014 mit der Universität Bielefeld nur eine Forschungsinstitution in Deutschland, die eine dedizierten FD-Richtlinie verabschiedet hat [15, S. 6]. Bis Ende 2017 wuchs diese Anzahl jedoch auf 22 institutionelle FD-Richtlinien in Deutschland an [37]. In anderen Ländern, wie in den Vereinigten Staaten oder dem Vereinigten Königreich, zeichnet sich ein ähnlicher Trend ab, der allerdings bereits früher begonnen hat als in Deutschland [15, 38].

<p>Eine Eigenheit Deutschlands ist hierbei die Heterogenität der Namen der FD-Richtlinien, auch wenn diese prinzipiell als zweckgleich gedacht waren [39]. So wurden 50 % ($n = 11$) der FD-Richtlinien in 2018 als <i>Leitlinien</i>, 23 % ($n = 5$) als <i>Grundsätze</i>, 18 % ($n = 4$) als <i>Policies</i> und 9 % ($n = 2$) als <i>Richtlinien</i> bezeichnet [39].</p>	<p>Heterogenität der Namen</p>
<p>Der Sinn einer FD-Richtlinie ist es, die Kosten- und Ressourcensteuerung für FDM zu optimieren und die Einhaltung von Förderrichtlinien zu garantieren. Zudem helfen sie Universitäten, ihre personellen, organisatorischen und technischen Kapazitäten für das FDM zu planen. Weitere Vorteile sind die Erhöhung der Reputation der Universität, Transparenz für Hochschulangehörige und Vorteile bei der Einwerbung von Fördermitteln [37, 39]. Auch empfehlen sie häufig, nach Möglichkeit, die Veröffentlichung als <i>Open Data</i>. Derlei Richtlinien unterstützen dann auch die Idee, dass öffentlich finanzierte Forschung der Öffentlichkeit auch zugänglich sein sollte, was wiederum der Forschungscommunity und der allgemeinen Gesellschaft zugutekommt.</p>	<p>Zweck</p>
<p>Der Inhalt deutscher FD-Richtlinien lässt sich dabei in fünf Kategorien aufteilen, welcher allerdings vom Umfang her zwischen den verschiedenen Richtlinien variiert: die Präambel, der Geltungsbereich, die rechtlichen Aspekte, der Umgang mit FD sowie die jeweiligen Verantwortlichkeiten der Forschenden und der Institution [37].</p>	<p>Inhalt</p>
<p>Die Präambel beinhaltet hierbei typischerweise eine Aussage zu der Bedeutung von FD und dem dazugehörigen Ziel der Institution sowie ein Verweis auf die GWP [37].</p>	<p>Präambel</p>
<p>Der Geltungsbereich beinhaltet die verabschiedende Instanz, das Datum der Veröffentlichung, die Zielgruppe der Richtlinie und der Titel der Richtlinie selbst [37].</p>	<p>Geltungsbereich</p>
<p>Die rechtlichen Aspekte umfassen die ethischen und rechtlichen Vorgaben zu FDM, sowie Angaben zum Verhältnis zu Dritten (z.B. Forschungsförderer) und allgemeine Richtlinien zur Lizenzierung [37].</p>	<p>Recht</p>
<p>Im Umgang mit FD wird meist definiert, was FD und FDM sind sowie wo, wie und wie lange diese gesichert werden sollten. Dabei wird auch auf die GWP verwiesen und empfohlen, fachspezifische Standards zu etablieren oder zu nutzen, Datenmanagementplans (DMPs) zu erstellen und FD, nach Möglichkeit, als <i>Open Data</i> zu veröffentlichen [37].</p>	<p>Umgang mit FD</p>
<p>Unter Verantwortlichkeiten wird spezifiziert, wer der Hauptverantwortliche für die FD ist (zumeist der Leiter des Forschungsprojektes) und welche Aufgaben dieser zu erfüllen hat. Zusätzlich werden häufig auch Verantwortungen seitens der Institution kodifiziert, wie das Bereitstellen der technischen Ausstattung bzw. Infrastruktur und die Organisation von Schulungen bzw. Beratungen zum Thema FD / FDM [37].</p>	<p>Verantwortlichkeit</p>
<p>Für eine genauere Inhaltsdarstellung und die Eingliederung im Vergleich zu internationalen Empfehlungen wird auf existierende Literatur verwiesen [37].</p>	<p>Inhaltsdetails</p>

2.3 Forschungsdaten in Dissertationen

Was die empirische Landschaft zu FD in Dissertationen angeht, so gibt es hierzu bisher kaum Studien–und jene, die es gibt, sind zumeist sehr begrenzt und spezialisiert im Umfang. So behandelte z.B. 2018 eine Abschlussarbeit an der Humboldt-Universität

zu Berlin die FD für eine repräsentative Stichprobe der musikwissenschaftlichen und -pädagogischen Dissertationen aus dem Jahre 2015 [9]. Diese analysierte welche Arten von FD in derlei Dissertationen vorkommen und wie diese publiziert wurden. Das Ergebnis hierbei war, dass nur zwei der 45 untersuchten Dissertationen mit zusätzlichen Datenpublikationen verknüpft wurde–und eine der beiden Datenpublikationen wurde vom Autor nur als halbe Publikation gewertet, da diese nur aus einem ausgelagerten Textanhang eines Buches bestand. Dies würde einer FD-Rate von 3,33 % ($n = 1,5$) entsprechen (eigene Berechnung auf Basis von Daten aus [9]).

Appendix-
Studie

Im Jahr 2015 wurde eine weitere Studie zu geistes- und sozialwissenschaftlichen Dissertationen und ihren FD durchgeführt [10]. Die Studie untersuchte 780 Dissertationen der Universitäten von Lille und Ljubljana, die zwischen 1987 und 2015 veröffentlicht wurden und aus mindestens 15 verschiedenen Disziplinen stammten. Dabei wurde analysiert, welche FD im Anhang der Dissertationen, entweder als fester Bestandteil der Dissertation selbst oder als separate Datei bzw. als separates Dokument, enthalten waren. Von diesen Dissertationen waren 45 % ($n = 353$) digitaler und 55 % ($n = 427$) gedruckter Natur. Die Ergebnisse zeigten, dass 62 % ($n = 219 \pm 1$) aller digitalen und 49 % ($n = 209 \pm 1$) aller gedruckten Dissertationen den Appendix in das Hauptdokument integriert hatten.¹ Respektive 7 % ($n = 25 \pm 1$) und 16 % ($n = 68 \pm 1$) der digitalen und gedruckten Dissertationen hatten einen Appendix, der separat vom Hauptdokument war und weitere respektive 28 % ($n = 99 \pm 1$) und 49 % ($n = 209 \pm 1$) hatten keinerlei Appendix. Zusätzlich war die Publikationsart des Appendix von 3 % ($n = 11 \pm 1$) der digitalen Dissertationen nicht näher spezifizierbar. Die Mehrheit der Appendixe enthielten, mit einer jeweiligen Frequenz von $n > 100$, Fragebögen (welche, entgegen der DFG-Richtlinie [14], nicht als FD klassifiziert wurden [10, S. 14]), Interviews, Textproben und experimentelle Observationen. Das Dateiformat der meisten Anhänge war hierbei PDF–auch für Tabellen, Fotografien und andere Inhalte, für die es geeignetere und FAIRere Alternativen gegeben hätte [10, S. 14]. Ein statistisch signifikanter Unterschied zwischen gedruckten und digitalen Dissertationen wurde nicht gefunden. Es wurde befunden, dass nur sehr wenige Dissertationen adäquates FDM bezeugen und dass die Mehrheit der FD durch ihre Publikationsart, mangelnde Organisation sowie generelle Unvollständigkeit nur unter schweren Bedingungen für weitere Forschung direkt weiterverwendbar seien [10, S. 20f.].

Präskriptive
Dokumente

Darüber hinaus besteht die Forschung zu der Beziehung zwischen Dissertationen und FD hauptsächlich aus präskriptiven Richtlinien, wie z.B. das *eDissPlus*-Projekt [5, 6, 7] und die daraus resultierende Richtlinie der Deutschen Nationalbibliothek (DNB) für die Pflichtabgabe von FD aus Dissertationen [8]. Empirische Studien zu der Umsetzung dieser Pflichtabgabe wurden jedoch noch nicht durchgeführt. Studien zu der Durchsetzung ähnlich verpflichtender Richtlinien anderer Länder, wie z.B. Frankreich, zeigen allerdings, dass diese nicht konsequent durchgesetzt werden und auch nicht notwendigerweise zur Publikation FAIRerer Daten führen [vgl. 10].

¹Die zitierte Studie gab in den meisten Fällen nur relative Werte an. Für diese Fälle wurden die absoluten Werte mit einer Unsicherheit von ± 1 Dissertation berechnet und werden hier entsprechend gekennzeichnet.

Kapitel 3

Richtlinien zu Forschungsdaten aus deutschen Promotionsvorhaben

Dieses Kapitel behandelt die verschiedenen verwaltungsrechtlichen Dokumente wissenschaftlicher Institutionen, die ein Promotionsvorhaben in Bezug auf FDM entweder spezifisch oder auch nur allgemein betreffen. Es wird überprüft, inwiefern promotionsberechtigte Institutionen in Deutschland bereits derlei Richtlinien erlassen haben, in welcher Form diese existieren und welche Anforderungen diese stellen.

Thema des Kapitels

Hierfür wird in Abschnitt 3.1 aufgeführt, wie die zu untersuchenden Institutionen ausgewählt wurden, wie die Materialien der Institutionen ausgesucht wurden und mit welchen Methoden das gesammelte Material daraufhin ausgewertet wurde. In Abschnitt 3.2 werden die entsprechenden Ergebnisse der Materialauswertung dargestellt. Abschließend werden in Abschnitt 3.3 die dargestellten Ergebnisse evaluiert und diskutiert.

Aufbau des Kapitels

3.1 Material & Methoden

In diesem Abschnitt wird das zu untersuchende Material in Abschnitt 3.1.1 und die Methoden der Untersuchung in Abschnitt 3.1.2 dargestellt.

Aufbau des Abschnitts

3.1.1 Material

Da es in Deutschland hierzu keine offizielle und öffentlich zugängliche Liste aller Universitäten mit Promotionsrecht seitens des BMBF gibt, wird als Datengrundlage für dieses Kapitel die von der HRK geführte Liste aller wissenschaftlichen Institutionen aus dem tertiären Bildungsbereich in Deutschland [40] genutzt. Diese Liste wird tagesaktuell, basierend auf der Selbstauskunft aller involvierten Institutionen ($n = 428$), geführt. Sie kodifiziert, unter anderem, welche Institutionen das Promotions- und Habilitationsrecht führen und umfasst auch Institutionen, die nicht Mitglied der HRK sind. Sie besitzt einen *de facto* wenn auch nicht *de jure* Status als Datengrundlage für allgemeine Informationen zu wissenschaftlichen Institutionen in Deutschland. Für die tagesspezifische Version der Liste, die für diese Arbeit genutzt wurde, siehe [41].

Datengrundlage

Um die zentrale Forschungsfrage dieses Kapitels zu beantworten, wurde diese Liste auf nur jene Institutionen gefiltert, welche das Promotionsrecht besitzen ($n = 163$). Die resultierende Liste promotionsberechtigter Institutionen besteht aus Forschungsinstitutionen verschiedener Hochschultypen sowie unterschiedlicher Trägerschaften. Von den Hochschultypen her umfasst die Liste Universitäten, Fachhochschulen (FHs), Hochschulen für Angewandte Wissenschaften (HAWs), Künstlerische Hochschulen

Grundmengenbeschreibung

(KHs), eine Verwaltungshochschule (VH) sowie eine Hochschule eigenen Typs (HSeT). Von den Trägerschaften her umfasst die Liste öffentlich-rechtliche, private sowie kirchliche Institutionen. Alle Institutionen sind staatlich anerkannt. Die relative sowie absolute Distribution aller promotionsberechtigter Institutionen in Deutschland nach Hochschultyp und Trägerschaft ist in Tabelle 3.1 gegeben.

Tabelle 3.1: Die Verteilung aller promotionsberechtigter Institutionen in Deutschland nach *Hochschultyp* × *Trägerschaft* aufgegliedert. Absolute Werte in Klammern angegeben.

	Öffentlich-Rechtlich	Privat	Kirchlich	Summe
Universität	53,37 % (87)	7,98 % (13)	6,13 % (10)	67,48 % (110)
FH / HAW	6,75 % (11)	0,00 % (0)	0,00 % (0)	6,75 % (11)
KH	23,93 % (39)	0,61 % (1)	0,00 % (0)	24,54 % (40)
HSeT	0,61 % (1)	0,00 % (0)	0,00 % (0)	0,61 % (1)
VH	0,61 % (1)	0,00 % (0)	0,00 % (0)	0,61 % (1)
Summe	85,28 % (139)	8,59 % (14)	6,13 % (10)	100,00 % (163)

Geografisch gesehen sind, zu unterschiedlich hohen Anteilen, Institutionen aus allen deutschen Bundesländern in der gefilterten Liste vertreten. Die genaue Verteilung ist in Abbildung 3.1 wiedergegeben, um diese mit der späteren Verteilung der Stichprobe vergleichen zu können.

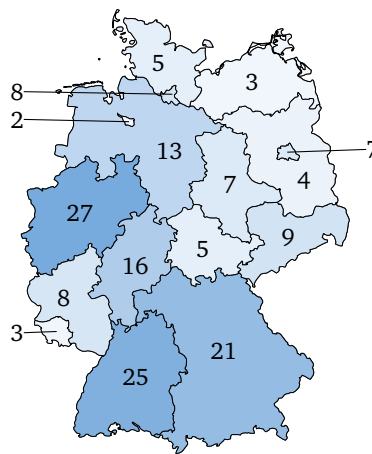


Abbildung 3.1: Die absolute Anzahl promotionsberechtigter Institutionen nach Bundesland. Steigende Farbintensivität korreliert mit steigender absoluten Anzahl.

Stichproben-
ziehung

Diese Liste promotionsberechtigter Institutionen bildete die Grundmenge für die Ziehung einer einfachen Zufallsstichprobe. Bei der Auswahl der Stichprobe wurde ein Konfidenzintervall von 95 % und eine Fehlerspanne von 5 % zugrunde gelegt. Diese Parameter gewährleisten, dass die Ergebnisse der Stichprobe mit hoher Wahr-

scheinlichkeit repräsentativ für die gesamte Population sind und die Unsicherheit der Schätzungen innerhalb akzeptabler Grenzen bleibt. Um den Prozess der Stichprobenziehung zu automatisieren und eine zufällige Auswahl zu gewährleisten, wurde eine auf Python basierende Software [42] genutzt, welche im Rahmen dieser Arbeit geschrieben wurde.¹

Die so gezogene Stichprobe ($n = 115$) besteht aus ca. 71 % aller promotionsberechtigter Institutionen. Die Stichprobe umfasst Institutionen aller Trägerschaften aus der Grundmenge: öffentlich-rechtliche, private sowie kirchliche Institutionen. Darüber hinaus umfasst die Stichprobe von den Hochschultypen her Universitäten, FHs, HAWs und eine HSeT. In der Stichprobe befindet sich nicht die VH, die sich in der Grundmenge befindet. Mit dieser Ausnahme sind somit alle anderen Hochschultypen vertreten. Die relative sowie die absolute Distribution aller Institutionen in der Stichprobe nach Hochschultyp und Trägerschaft ist in Tabelle 3.2 gegeben.

Stichproben-
beschreibung

Tabelle 3.2: Die Verteilung der Institutionen in der Stichprobe nach *Hochschultyp* \times *Trägerschaft* aufgegliedert. Absolute Werte in Klammern angegeben.

	Öffentlich-Rechtlich		Privat		Kirchlich		Summe	
Universität	55,65 %	(64)	8,70 %	(10)	4,35 %	(5)	68,70 %	(79)
FH / HAW	6,96 %	(8)	0,00 %	(0)	0,00 %	(0)	6,96 %	(8)
KH	22,61 %	(26)	0,87 %	(1)	0,00 %	(0)	23,48 %	(27)
HSeT	0,87 %	(1)	0,00 %	(0)	0,00 %	(0)	0,87 %	(1)
VH	0,00 %	(0)	0,00 %	(0)	0,00 %	(0)	0,00 %	(0)
Summe	86,09 %	(99)	9,57 %	(11)	4,35 %	(5)	100,00 %	(115)

Geografisch gesehen sind, zu unterschiedlich hohen Anteilen, Institutionen aus allen deutschen Bundesländern in der Stichprobe vertreten. Die genaue Verteilung ist in Abbildung 3.2 wiedergegeben.

Für die Evaluation, inwiefern die Institutionen der Stichprobe verwaltungsrechtliche Dokumente besitzen, die entweder allgemeine FDM-Richtlinien für alle Forschenden der Institution oder spezifische Regelungen für Promovierende beinhalten, wurde deren gesamte öffentlich zugängliche Online-Präsenz nach relevanten Dateien und Webseiten durchsucht. Diese Suche fand im Allgemeinen über zwei Wege statt: die interne Suchfunktion der Institution und die externe Durchsuchung der Institutions-Domäne via der Suchmaschine *DuckDuckGo*.

Dokumente-
sammlung

Für die allgemeinen Richtlinien wurden in erster Linie eigenständige FDM-Richtlinien gesucht. Hierfür wurde, jenseits der bereits beschriebenen allgemeinen Suchmethode auch die dedizierte Forschungsdatenpolicies-Liste der Informationsplattform

Allgemeine
Dokumente

¹Die Software von KRASSNIG [42] nutzt standardmäßig die Anzahl an Nanosekunden seit dem Beginn der System-Epoche (1970-01-01T00:00:00Z) als Startwert für die Zufallsfunktion. Der genutzte Startwert wird als begleitendes Metadatum der Stichprobe abgespeichert. Die Ziehung ist somit wiederholbar und das Datum der Ziehung verifizierbar.

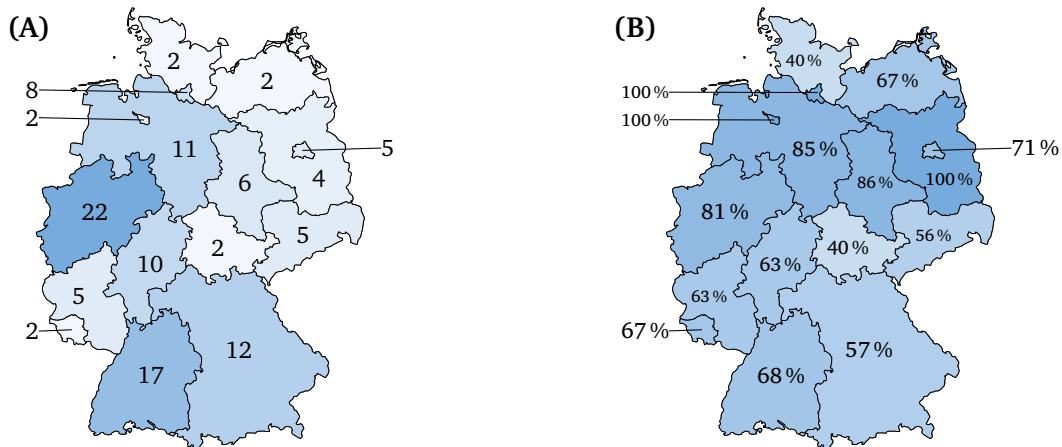


Abbildung 3.2: Verteilung der Stichproben-Institutionen nach Bundesland. (A) Die absolute Anzahl der Institutionen nach Bundesland. (B) Der Anteil der Institutionen, die aus der Grundmenge übernommen wurden, relativ zum jeweiligen Bundesland. Steigende Farbintensivität zeigt relative Nähe zum Maximalwert an.

FORSCHUNGSDATEN.ORG genutzt [43].² Jenseits der spezifischen FDM-Richtlinien wurden auch Richtlinien zur Sicherung der GWP gesucht, da auch diese den Umgang mit FD betreffen. Zusätzlich wurden auch andere Richtlinien gesucht, die den Forschenden Regeln zu FD oder FDM auferlegen, aber weder dedizierte FD-Richtlinien noch Regeln zum Einhalten der GWK darstellen (diese Kategorie wird fortan als *Andere Richtlinien* bezeichnet). Wenn es von einem Dokument sowohl eine HTML- wie auch eine PDF-Datei gab, so wurde nur die PDF-Datei zur Evaluation weitergenutzt. Hierbei wurden insgesamt 142 Dokumente zur Weiterverarbeitung aufgenommen.

Promotions-
spezifische
Dokumente

Für promotionsspezifische Richtlinien wurden Promotions- und Prüfungsordnungen gesucht. Dabei wurden sowohl fachspezifische Ordnungen wie auch verbindliche Rahmenbedingungen und anderweitige übergreifende Ordnungen aufgenommen. Die heterogene Handhabung dieser Dokumente seitens der Institutionen führte zu folgenden Selektionsregeln bei der Auswahl: (i) Wenn es eine aktuelle Lesefassung der Promotionsordnung gibt, so wird diese bevorzugt. (ii) Sollte es keine aktuelle Lesefassung der Promotionsordnung geben, so wird die aktuellste Gesamtversion der Promotionsordnung bevorzugt. (iii) Sollte es keine aktuelle Gesamtversion geben, so werden, zusätzlich zu der letzten Version der Promotionsordnung, auch alle seither erschienen relevanten verwaltungsrechtlichen Addenda aufgenommen. (iv) Wenn es von einem Dokument sowohl eine HTML- wie auch eine PDF-Datei gab, so wurde nur die PDF-Datei zur Evaluation weitergenutzt. Hierbei wurden insgesamt 754 Dokumente zur Weiterverarbeitung aufgenommen.

²Die FDM-Richtlinien, welche sich nicht auf diesem Portal haben finden lassen, werden vom Autor nach Beendigung dieser Arbeit dort nach Möglichkeit nachgetragen.

3.1.2 Methoden

Die in Abschnitt 3.1.1 gesammelten allgemeinen Dokumente und deren Institutionen wurden dann wie folgt klassifiziert: (i) Jedes Dokument wurde durch manuelle Überprüfung einem Typ zugeordnet. Diese Typen waren, hierarchisch geordnet:

*Allgemeine
Dokumente*

1. Nicht relevant
2. Richtlinie zu GWP
3. Anderweitige Richtlinie die FDM beinhaltet
4. Richtlinie zu FD / FDM

(ii) Bei Dokumenten, welche als „Richtlinie zu FD / FDM“ klassifiziert wurden, wird zusätzlich notiert, ob es sich dabei um eine Leitlinie, einen Grundsatz, eine Policy, eine Empfehlung oder eine Richtlinie handelt (nach dokumenteigener Angabe). (iii) Jede Institution erhielt dann die Klassifikation des Dokumentes, welche die höchste hierarchische Klassifikationsstufe besitzt, es sei denn, die Institution hatte kein öffentlich zugängliches Dokument dieser Art. In diesem Fall wurde dies stattdessen als Klassifikationsstufe vermerkt.

Die in Abschnitt 3.1.1 gesammelten promotionsspezifischen Dokumente und deren Institutionen wurden dann wie folgt klassifiziert: (i) Es wurde überprüft, ob alle PDF-Dateien lesbaren eingebetteten Text besitzen. Hierfür wurde mit einem Bash-Skript [41] der Anfang und das Ende des eingebetteten Textes via *pdftotext* aus der *Poppler*-Softwaresammlung [44] angezeigt. (ii) Texte, die sich als maschinell nicht lesbar erwiesen haben, wurden notiert und später manuell klassifiziert. (iii) Die restlichen Texte wurden zuerst mit einem Bash-Skript [41] darauf überprüft, ob sie Text beinhalten, der FD, FDM oder GWP betrifft. (iv) Die durch das Skript angezeigten Treffer wurden dann auf Kontext überprüft und entsprechend manuell klassifiziert. Hierbei gab es fünf Klassifikationsstufen:

*Promotions-
spezifische
Dokumente*

1. Keinerlei Richtlinien zu FD / FDM enthalten
2. Richtlinien zu GWP als Empfehlung enthalten
3. Richtlinien zu GWP als Verpflichtung enthalten
4. Richtlinien zu FD / FDM als Empfehlung enthalten
5. Richtlinien zu FD / FDM als Verpflichtung enthalten

(v) Jede Institution erhielt dann die Klassifikation des Dokumentes, welche die höchste hierarchische Klassifikationsstufe besitzt, es sei denn, die Institution hatte kein öffentlich zugängliches Dokument dieser Art. In diesem Fall wurde dies stattdessen als Klassifikationsstufe vermerkt. Es wurde zusätzlich notiert, ob die promotionsspezifischen Richtlinien für alle Promovierenden gelten oder ob diese nur eine Teilmenge der Promovierenden betreffen.

Statistische
Auswertung

Für die Kreuzprodukte aller vermuteter Faktoren sowie für die Klassifikation beider Dokumentgruppen wurden Chi-Quadrat-Tests für Unabhängigkeit durchgeführt, um zu überprüfen, ob statistisch signifikante Relationen zwischen der Dokumentklassifikationsverteilung, den Institutionstypen, den Bundesländern und den Trägerschaften besteht. Hierbei wurde für alle zu überprüfenden Relationen von der Nullhypothese ausgegangen. Für die Auswertung der Richtlinienklassifikation wurde dabei davon ausgegangen, dass Bundesland, Trägerschaft und Institutionstyp jeweils als unabhängige Variable fungieren und die Klassifikationsstufe der Institutionen als abhängige Variable. Die Nullhypothese sagt vorher, dass die Werte der unabhängigen Variable keinen Einfluss auf die Werte der abhängigen Variable haben werden: I.e. für die Kombination *Faktor A* \times *Faktor B* wird angenommen, dass *Faktor A* keinen Einfluss auf *Faktor B* hat und dass, bedingt durch die symmetrische Natur des Chi-Quadrat-Tests, auch andersherum kein Einfluss besteht. Die Nullhypothese gilt als widerlegt wenn der respektive Chi-Quadrat-Test für Unabhängigkeit einen Signifikanzwert von $p < 0,05$ erzeugt. Bei Signifikanzwerten von $p \geq 0,05$ gilt die Nullhypothese als bestätigt [45, 46]. Da p-Werte jedoch nichts über die Stärke einer Abhängigkeit aussagen, wurden für alle Testergebnisse mit $p < 0,05$ zusätzlich der respektive Cramérs V-Wert (ϕ_C) berechnet, um zu überprüfen, wie stark die statistisch signifikante Abhängigkeit im Endeffekt ist [47]. Bei einem Cramérs V-Wert von $\phi_C > 0,1$ ist von einem schwachem, bei $\phi_C > 0,3$ von einem moderaten und bei $\phi_C > 0,5$ von einem starken Zusammenhang bzw. Einfluss auszugehen.

Software

Alle statistischen Auswertungen wurden via R ausgeführt [41, 48, 49]. Für die Klassifikationsarbeit sowie für alle anderweitigen manuellen Bearbeitungen der Datentabellen wurde *LibreOffice Calc* genutzt [50].

3.2 Resultate

Aufbau des
Abschnitts

In diesem Abschnitt werden die Resultate der Klassifizierung sowie deren statistische Auswertung dargestellt. In Abschnitt 3.2.2 werden die Resultate und Auswertung der allgemeinen verwaltungsrechtlichen Dokumente aufgelistet. In Abschnitt 3.2.3 werden die Resultate und Auswertung der promotionsspezifischen Dokumente aufgelistet.

3.2.1 Mögliche relevante Faktoren

Jenseits der bereits in Abschnitt 3.1 genannten Faktoren, konnten keine weiteren potenziell relevanten Faktoren identifiziert werden.

Unabhängig-
keit der
Faktoren

Zur Überprüfung, ob die zu untersuchenden Faktoren voneinander abhängig sind, wurden für die Kreuzprodukte aller Faktorenkombinationen–Bundesland, Institutionstyp und Trägerschaft–Chi-Quadrat Tests der Unabhängigkeit durchgeführt. Die Interaktionen *Bundesland* \times *Institutionstyp* ($\chi^2(45, n = 115) = 52,84, p = 0,20 > 0,05$), *Bundesland* \times *Trägerschaft* ($\chi^2(30, n = 115) = 22,18, p = 0,85 > 0,05$) und *Institutionstyp* \times *Trägerschaft* ($\chi^2(6, n = 115) = 5,66, p = 0,46 > 0,05$) waren alle statistisch nicht signifikant und gelten daher als voneinander unabhängig.

3.2.2 Allgemeine Dokumente

Die in Abschnitt 3.1.1 gesammelten 142 Dokumente allgemeiner verwaltungsrechtlicher Natur wurden, wie in Abschnitt 3.1.2 beschrieben, manuell ausgewertet und in vier verschiedene Klassifikationsstufen aufgeteilt. Das Dokument mit der höchsten Klassifikationsstufe seiner Institution wurde dann als Klassifikationsstufe der Institution verwendet.

Methodenzusammenfassung

Von den 115 Institutionen der Stichprobe besitzen 44,35 % ($n = 51$) eine dedizierte FD-Richtlinie für alle Forschenden. Weitere 43,48 % ($n = 50$) der Institutionen besitzen nur eine Richtlinie zu GWP. Es gab keine Institutionen, die ein Dokument besitzen, was Richtlinien zu FD jenseits der GWP enthält, die nicht auch eine dedizierte FD-Richtlinie besitzen. Die restlichen 12,17 % ($n = 14$) der Institutionen hatten keine der oben genannten Richtlinien.

Allgemeine Auswertung

Nach Institutionstyp aufgegliedert, ergibt sich hierbei die Verteilung aus Tabelle 3.3. Der Chi-Quadrat-Test für *Institutionstyp* \times *Richtlinienklassifikation* ergab einen Wert

Institutionstyp

Tabelle 3.3: Verteilung der höchsten Klassifizierungsstufe der Institutionen durch ihre allgemeingültigen Dokumente nach *Institutionstyp* \times *Klassifikationsstufe* aufgegliedert. Angabe relativ zum jeweiligen Institutionstyp. Absolute Werte in Klammern angegeben.

	Keine Verfügbar		GWP-Richtlinien		Andere Richtlinien		FD-Richtlinien	
Universität	5,06 %	(4)	35,44 %	(13)	0,00 %	(0)	59,49 %	(47)
FH / HAW	0,00 %	(0)	75,00 %	(6)	0,00 %	(0)	25,00 %	(2)
KH	37,04 %	(10)	55,56 %	(15)	0,00 %	(0)	7,41 %	(2)
HSeT	0,00 %	(0)	100,00 %	(1)	0,00 %	(0)	0,00 %	(0)
Alle	12,17 %	(14)	43,48 %	(50)	0,00 %	(0)	44,35 %	(51)

von $\chi^2(6, n = 115) = 36,24$, $p = 2,47 \cdot 10^{-6} < 0,001$ und einen dazugehörigen Cramérs V-Wert von $\phi_C = 0,40$. Die Interaktion ist somit statistisch hochsignifikant und besitzt eine moderate Effektstärke.

Nach Trägerschaft aufgegliedert, ergibt sich hierbei die in Tabelle 3.4 dargestellte Verteilung. Die Interaktion zwischen *Trägerschaft* und *Allgemeine Richtlinien* ist mit

Trägerschaft

Tabelle 3.4: Verteilung der höchsten Klassifizierungsstufe der Institutionen durch ihre allgemeingültigen Dokumente nach *Trägerschaft* \times *Klassifikationsstufe* aufgegliedert. Angaben relativ zur jeweiligen Trägerschaft. Absolute Werte in Klammern angegeben.

	Keine Verfügbar		GWP-Richtlinien		Andere Richtlinien		FD-Richtlinien	
Öffentlich-Rechtlich	10,10 %	(10)	39,39 %	(39)	0,00 %	(0)	50,51 %	(50)
Privat	18,18 %	(2)	72,73 %	(8)	0,00 %	(0)	9,09 %	(1)
Kirchlich	40,00 %	(2)	60,00 %	(3)	0,00 %	(0)	0,00 %	(0)
Alle	12,17 %	(14)	43,48 %	(50)	0,00 %	(0)	44,35 %	(51)

$\chi^2(4, n = 115) = 12,86, p = 0,01 < 0,05, \phi_C = 0,24$ statistisch signifikant und von schwacher Effektstärke.

Bundesländer

Nach *Bundesland* aufgegliedert, ergibt sich die in Abbildung 3.3 dargestellte Verteilung. Für eine tabellarische Darstellung der in Abbildung 3.3 dargestellten Informatio-

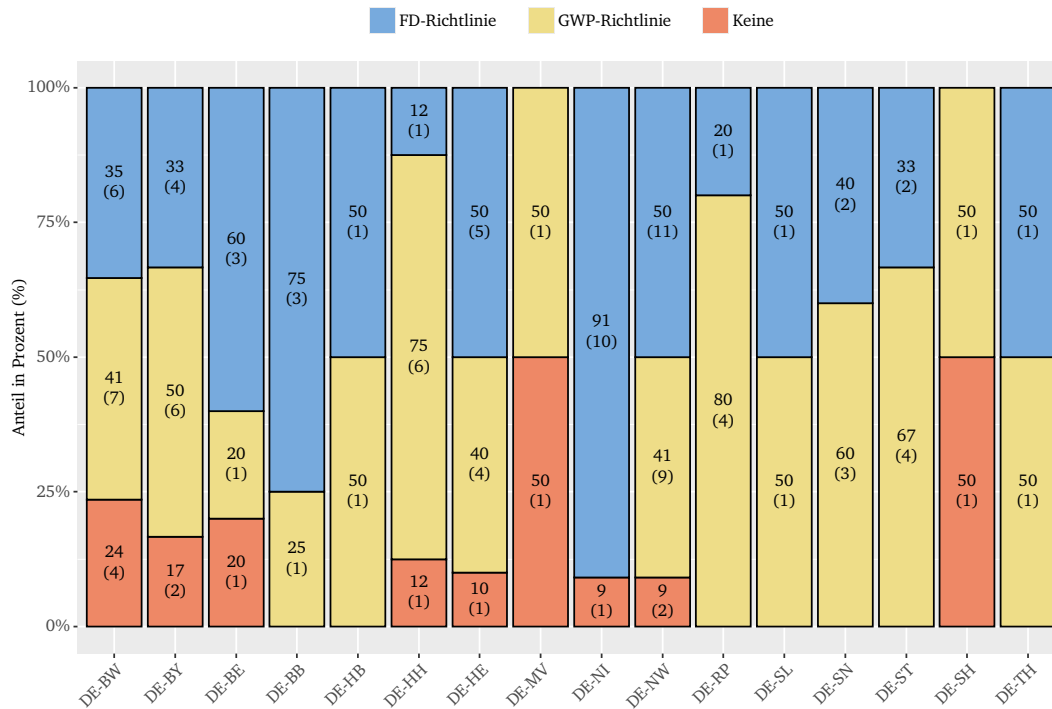


Abbildung 3.3: Verteilung der höchsten Klassifizierungsstufe der Institutionen durch ihre allgemeingültigen Dokumente nach *Bundesland* × *Klassifikationsstufe* aufgegliedert. Angaben relativ zum jeweiligen Bundesland. Absolute Werte in Klammern angegeben. Angabe der Bundesländer nach ISO 3166-2:DE [51].

nen, siehe Tabelle A.1. Der Chi-Quadrat-Test für *Bundesland* × *Richtlinienklassifikation* ergab einen Wert von $\chi^2(30, n = 115) = 32,74, p = 0,3341388 > 0,05$. Die Interaktion ist somit nicht signifikant.

Bezeichnung und Inhalt der Forschungsdaten-Dokumente

Bezeichnung

Es wurde auch überprüft, wie die einzelnen Institutionen ihre FD-Richtlinien offiziell nennen. Hierbei gab es insgesamt sieben verschiedene Bezeichnungen: *Empfehlungen*, *Leitlinie*, *Strategiepapier*, *Policy*, *Leitfaden*, *Grundsätze* und *Richtlinie*. Ein Anteil der Institutionen nutzte dabei eine Mehrfachbezeichnung von sowohl *Policy* wie auch *Richtlinie*. Die relative und absolute Verteilung dieser unterschiedlichen Namensgebung für FD-Richtlinien ist in Abbildung 3.4 gegeben.

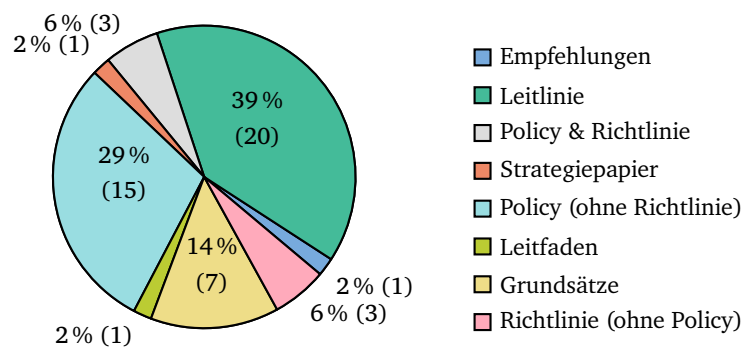


Abbildung 3.4: Anteil der verschiedenen Bezeichnungen für FD-Richtlinien. Absolute Werte in Klammern angegeben.

Inhaltlich ist die Mehrheit der Dokumente zum größten Teil deckungsgleich: So verweisen die Richtlinien in der Regel auf die FAIR-Prinzipien, auf die Regeln der GWP und auf die Angebote und Kapazitäten der eigenen Universität. Von den 51 Institutionen mit expliziten FD-Richtlinien hatten 78,43 % ($n = 40$) einen entweder verbindlichen oder zumindest scheinbar verbindlichen Charakter für Institution und Forschende.³ In den verbindlichen Richtlinien war zusätzlich zu dem oben angegebenen Inhalt auch meistens eine Klärung der Verantwortlichkeit zwischen der Institution und den Forschenden in Bezug auf die verschiedenen Stadien und Tätigkeiten im FDM vorhanden. Die restlichen 21,57 % ($n = 11$) gliedern sich in jene Richtlinien auf, die entweder nicht öffentlich zugänglich waren oder nur einen empfehlenden bzw. fördernden Charakter haben. Einen empfehlenden oder fördernden Charakter hatten 19,61 % ($n = 10$). Diese Dokumente beinhalteten meistens eine Bekennung der Universität, dass korrektes FDM wichtig ist und dass das Einhalten der FAIR-Regeln und/oder nach Möglichkeit eine Veröffentlichung als Open Data empfohlen wird. Nicht zugänglich waren 1,96 % ($n = 1$) der Dokumente.

Inhalt

Nach Dokumentname aufgegliedert, ergibt sich hierbei die in Tabelle 3.5 dargestellte Verteilung für Dokumentcharakter.

Klassifizierung

Ein Chi-Quadrat-Test auf Unabhängigkeit wurde durchgeführt, um zu überprüfen, ob es eine signifikante Relation zwischen dem Namen einer FD-Richtlinie und dessen inhaltlichen Charakter (verpflichtend oder empfehlend bzw. fördernd) gibt. Die *Dokumentname* × *Charakter*-Interaktion ergab einen Wert von $\chi^2(24, n = 51) = 63,34, p = 2,12 \cdot 10^{-5} < 0,001, \phi_C = 0,64 > 0,5$ und war somit sowohl statistisch hoch signifikant wie auch vom Effekt her stark.

Signifikanz

³Die genaue Aufteilung zwischen verbindlich und scheinbar verbindlich ist durch den Verweis auf externe Dokumente und wegen legalistischer Sprache schwer im Rahmen dieser Arbeit zu unterscheiden. Diese werden daher zusammen gruppiert.

Tabelle 3.5: Dokumentcharakter der FD-Richtlinien in relativer Angabe zum Dokumentnamen. Absolute Werte in Klammern angegeben.

	Pflicht		Empfehlung		Nicht zugänglich	
Empfehlungen	0,00 %	(0)	100,00 %	(1)	0,00 %	(0)
Grundsätze	85,71 %	(6)	14,29 %	(1)	0,00 %	(0)
Leitfaden	100,00 %	(1)	0,00 %	(0)	0,00 %	(0)
Leitlinie	70,00 %	(14)	30,00 %	(6)	0,00 %	(0)
Policy	86,67 %	(13)	13,33 %	(2)	0,00 %	(0)
Policy & Richtlinie	100,00 %	(3)	0,00 %	(0)	0,00 %	(0)
Richtlinie	100,00 %	(3)	0,00 %	(0)	0,00 %	(0)
Strategiepapier	0,00 %	(0)	0,00 %	(0)	100,00 %	(1)
Alle	78,43 %	(40)	19,61 %	(10)	1,96 %	(1)

3.2.3 Promotionspezifische Dokumente

Methodenzusammenfassung

Die in Abschnitt 3.1.1 gesammelten 754 Dokumente promotionspezifischer Natur wurden, wie in Abschnitt 3.1.2 beschrieben, unter Beihilfe maschineller Filter manuell ausgewertet und in fünf verschiedene Klassifikationsstufen aufgeteilt. Die Anzahl der Klassifikationsstufen erhöht sich auf sechs, wenn inkludiert wird, dass manche Institutionen ihre Promotions- und Prüfungsordnungen nicht öffentlich zugänglich gemacht haben. Das Dokument mit der höchsten Klassifikationsstufe seiner Institution wurde dann als Klassifikationsstufe der Institution verwendet.

Allgemeine Auswertung

Von den 115 Institutionen der Stichprobe machten 3,48 % ($n = 4$) keine ihrer Prüfungs- und Promotionsordnungen öffentlich zugänglich. Diese Dokumente und dazugehörigen Institutionen konnten daher nicht weiter evaluiert werden. Weitere 25,22 % ($n = 29$) erwähnen weder Richtlinien zu GWP noch Richtlinien zu FD in ihren Promotions- bzw. Prüfungsordnungen. Aus der Stichprobe hatten insgesamt 56,52 % ($n = 65$) der Institutionen mindestens eine Prüfungs- oder Promotionsordnung, die die Regeln der GWP erwähnen. Aus diesen Institutionen hatten 6,15 % ($n = 4$) höchstens eine nicht explizit verbindliche Erwähnung der GWP, während die restlichen 93,85 % ($n = 61$) dieser Institutionen sie für Promovierende explizit verbindlich gültig gemacht haben. In Relation zu der gesamten Stichprobe entsprechen diese Institutionen mit GWP-Richtlinien jeweils 3,48 % ($n = 4$) und 53,04 % ($n = 61$). Zuletzt hatten die restlichen 14,78 % ($n = 17$) der Institutionen mindestens eine spezifische Richtlinie zum Umgang mit bzw. der Veröffentlichung von FD, die im Rahmen des Promotionsvorhabens entstanden sind. Aus dieser Menge hatten 5,88 % ($n = 1$) höchstens eine nicht explizit verbindliche Erwähnung zum Umgang mit FD, während die restlichen 94,12 % ($n = 16$) FD-Richtlinien verbindlicher Natur beinhalteten. Diese Institutionen mit FD-Richtlinien machen jeweils 0,87 % ($n = 1$) und 13,91 % ($n = 16$) der gesamten Stichprobe aus.

Nach Institutionstyp aufgegliedert, ergibt sich hierbei die Verteilung in Tabelle 3.6. Institutionstyp

Tabelle 3.6: Verteilung der höchsten Klassifizierungsstufe der Institutionen durch ihre promotionsspezifische Dokumente nach *Institutionstyp* × *Richtlinienklassifikation* aufgegliedert. Angaben relativ zum jeweiligen Institutionstyp. Absolute Werte in Klammern angegeben.

	Kein(e)		GWP-Richtlinie		FD-Richtlinie	
	Zugang	FD-Richtlinie	Empfehlung	Verpflichtung	Empfehlung	Verpflichtung
Universität	2,53 % (2)	13,92 % (11)	2,53 % (2)	64,56 % (51)	1,27 % (1)	15,19 % (12)
FH / HAW	0,00 % (0)	12,50 % (1)	0,00 % (0)	37,50 % (3)	0,00 % (0)	50,00 % (4)
KH	7,41 % (2)	62,96 % (17)	7,41 % (2)	22,22 % (6)	0,00 % (0)	0,00 % (0)
HSeT	0,00 % (0)	0,00 % (0)	0,00 % (0)	100,00 % (1)	0,00 % (0)	0,00 % (0)
Alle	3,48 % (4)	25,22 % (29)	3,48 % (4)	53,04 % (61)	0,87 % (1)	13,91 % (16)

Die Interaktion *Institutionstyp* × *Richtlinienklassifikation* war mit $\chi^2(15, n = 115) = 42,97$, $p = 1,59 \cdot 10^{-4} < 0,001$, $\phi_C = 0,35 > 0,3$ statistisch hochsignifikant und von moderater Effektstärke.

Nach Trägerschaft aufgegliedert, ergibt sich die in Tabelle 3.7 dargestellte Verteilung. Trägerschaft

Tabelle 3.7: Verteilung der höchsten Klassifizierungsstufe der Institutionen durch ihre promotionsspezifische Dokumente nach *Trägerschaft* × *Richtlinienklassifikation* aufgegliedert. Angaben relativ zur jeweiligen Trägerschaft. Absolute Werte in Klammern angegeben.

	Kein(e)		GWP-Richtlinie		FD-Richtlinie	
	Zugang	FD-Richtlinie	Empfehlung	Verpflichtung	Empfehlung	Verpflichtung
Öffentlich-Rechtlich	2,02 % (2)	23,23 % (23)	4,04 % (4)	55,56 % (55)	1,01 % (1)	14,14 % (14)
Privat	18,18 % (2)	27,27 % (3)	0,00 % (0)	36,36 % (4)	0,00 % (0)	18,18 % (2)
Kirchlich	0,00 % (0)	60,00 % (3)	0,00 % (0)	40,00 % (2)	0,00 % (0)	0,00 % (0)
Alle	3,48 % (4)	25,22 % (29)	3,48 % (4)	53,04 % (61)	0,87 % (1)	13,91 % (16)

Auch hier wurde ein Chi-Quadrat-Test auf Unabhängigkeit durchgeführt, um zu überprüfen, ob es eine signifikante Relation zwischen *Institutionstyp* und *Richtlinienklassifikation* gibt. Mit $\chi^2(10, n = 115) = 12,69$, $p = 0,24$ war dies nicht der Fall.

Nach Bundesländer aufgegliedert, ergibt sich die in Abbildung 3.5 dargestellte Verteilung. Für eine tabellarischer Darstellung der in Abbildung 3.5 aufgelisteten Informationen, siehe Tabelle A.2. Für *Bundesland* × *Richtlinienklassifikation* wurde mit $\chi^2(75, n = 115) = 103,66$, $p = 0,02$, $\phi_C = 0,42$ ein statistisch signifikanter und moderater Effekt gefunden. Bundesländer

Zusätzlich zu beachten sind noch etwaige Interaktionen zwischen den Institutsklassifikationsstufen für allgemeine und promotionsspezifische Dokumente. Die Verteilung aller Institutionen nach *Allgemeine Dokumentklassifikation* × *Spezifische Dokumentklassifikation* aufgegliedert befindet sich hierfür in Tabelle 3.8. Allgemeine und spezifische Dokumente

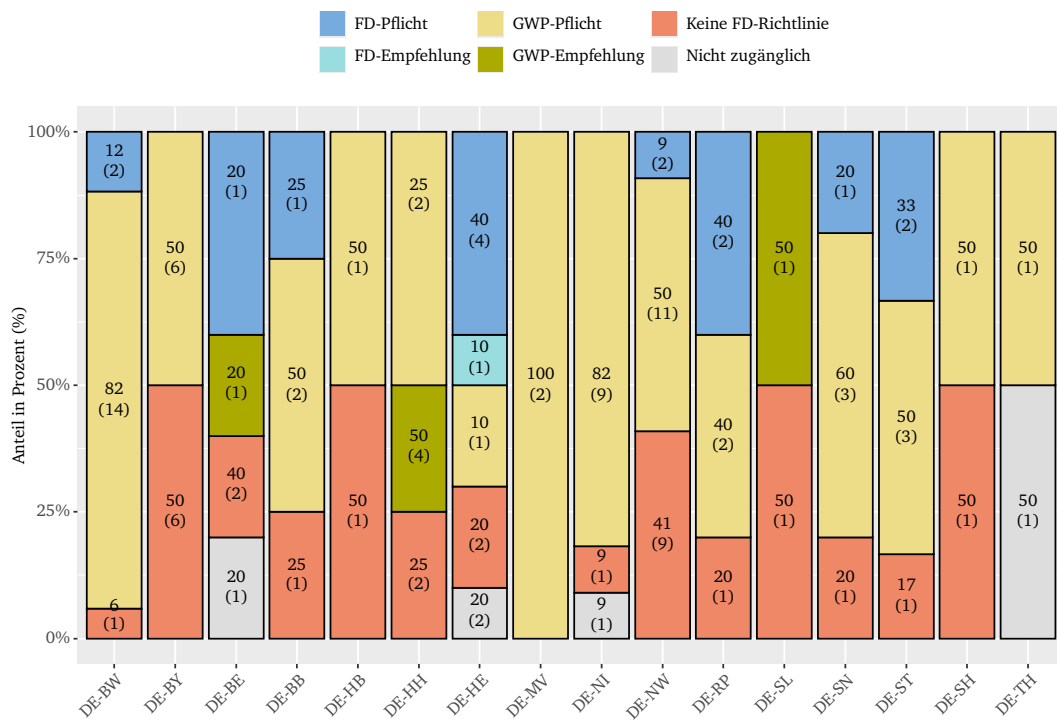


Abbildung 3.5: Verteilung der höchsten Klassifizierungsstufe der Institutionen durch ihre promotionsspezifische Dokumente in relativer Angabe zum Bundesland. Absolute Werte in Klammern angegeben. Angabe der Bundesländer nach ISO 3166-2:2020 [51].

Tabelle 3.8: Verteilung der höchsten Klassifizierungsstufe der Institutionen durch ihre promotionsspezifische Dokumente nach Allgemeine Dokumentklassifikation × Spezifische Dokumentklassifikation aufgegliedert. Angaben relativ zum Gesamtwert der Stichprobeninstitutionen. Absolute Werte in Klammern angegeben.

	Kein(e)		GWP-Richtlinie			FD-Richtlinie		
	Zugang	FD-Richtlinie	Empfehlung	Verpflichtung	Empfehlung	Verpflichtung		
Keine	1,74 % (2)	5,22 % (6)	0,00 % (0)	5,22 % (6)	0,00 % (0)	0,00 % (0)		
GWP	1,74 % (2)	15,65 % (18)	2,61 % (3)	20,00 % (23)	0,00 % (0)	3,48 % (4)		
FD	0,00 % (0)	4,35 % (5)	0,87 % (1)	27,83 % (32)	0,87 % (1)	10,43 % (12)		
Alle	3,48 % (4)	25,22 % (29)	3,48 % (4)	53,04 % (61)	0,87 % (1)	13,91 % (16)		

Die höchsten Klassifikationsstufen der beiden Dokumentgruppen stehen jeweils wie folgt in Relation zu einander: Institutionen mit Promotions- oder Prüfungsordnungen, die verpflichtenden Inhalt zu FD haben, waren zu 75,00 % auch Institutionen, die eine allgemeine verpflichtende FD-Richtlinie besitzen. In umgekehrter Relation hingegen, waren nur 23,53 % der Institutionen mit FD-Richtlinien auch Institutionen, welche mindestens eine Promotions- oder Prüfungsordnung besitzen, die den Umgang mit FDM verbindlich festlegt. Die gesamten relativen Anteile nach *Allgemeine Dokumentklassifikation* und nach *Spezifische Dokumentklassifikation* befinden sich respektiv in Tabelle 3.9 und Tabelle 3.10.

Tabelle 3.9: Die Verteilung der Institutionen nach *Allgemeine Dokumentklassifikation* × *Spezifische Dokumentklassifikation* aufgegliedert. Angaben relativ zu *Allgemeine Dokumentklassifikation*. Absolute Werte in Klammern angegeben.

	Kein(e)		GWP-Richtlinie				FD-Richtlinie			
	Zugang	FD-Richtlinie	Empfehlung	Verpflichtung	Empfehlung	Verpflichtung	Empfehlung	Verpflichtung		
Keine	14,29 % (2)	42,86 % (6)	0,00 % (0)	42,86 % (6)	0,00 % (0)	0,00 % (0)	0,00 % (0)	0,00 % (0)		
GWP	4,00 % (2)	36,00 % (18)	6,00 % (3)	46,00 % (23)	0,00 % (0)	8,00 % (4)				
FD	0,00 % (0)	9,80 % (5)	1,96 % (1)	62,75 % (32)	1,96 % (1)	23,53 % (12)				

Tabelle 3.10: Die Verteilung der Institutionen nach *Allgemeine Dokumentklassifikation* × *Spezifische Dokumentklassifikation* aufgegliedert. Angaben relativ zu *Spezifische Dokumentklassifikation*. Absolute Werte in Klammern angegeben.

	Kein(e)		GWP-Richtlinie				FD-Richtlinie			
	Zugang	FD-Richtlinie	Empfehlung	Verpflichtung	Empfehlung	Verpflichtung	Empfehlung	Verpflichtung		
Keine	50,00 % (2)	20,69 % (6)	0,00 % (0)	9,84 % (6)	0,00 % (0)	0,00 % (0)	0,00 % (0)	0,00 % (0)		
GWP	50,00 % (2)	62,07 % (18)	75,00 % (3)	37,70 % (23)	0,00 % (0)	25,00 % (4)				
FD	0,00 % (0)	17,24 % (5)	25,00 % (1)	52,46 % (32)	100,00 % (1)	75,00 % (12)				

Mit $\chi^2(10, n = 115) = 26,59, p = 0,003, \phi_C = 0,34$ war der Zusammenhang von *Allgemeine Dokumentklassifikation* × *Spezifische Dokumentklassifikation* statistisch signifikant und von moderater Effektstärke.

3.3 Diskussion

Die Auswertung der vermuteten Faktoren *Bundesland*, *Institutionstyp* und *Trägerschaft* in Abschnitt 3.2 zeigen, dass diese Faktoren von einander unabhängig sind, da die respektiven Chi-Quadrat-Teste p-Werte von $p > 0,05$ ergaben und damit die bestehenden Nullhypothesen bestätigten. Dementsprechend sind alle folgenden signifikanten Relationen zwischen diesen Faktoren und den zwei verschiedenen Dokumentklassifikationswerten als eigenständig signifikant betrachtbar.

Unabhängigkeit der Faktoren

3.3.1 Allgemeine Dokumente

Signifikante Faktoren Zwischen den oben genannten Faktoren und der höchsten Klassifikationsstufe einer Institution für allgemeine verwaltungsrechtliche Dokumente gab es folgende signifikante Relationen: Mit einem p-Wert von $p < 0,05$ waren die Abhängigkeiten zwischen *Institutionstyp* \times *Richtlinienklassifikation* und *Trägerschaft* \times *Richtlinienklassifikation* statistisch signifikant. Zusätzlich war die Abhängigkeit zwischen *Institutionstyp* und *Richtlinienklassifikation* mit einem p-Wert von $p < 0,001$ statistisch hochsignifikant. Diese zwei Chi-Quadrat-Tests haben dementsprechend die respektiven Nullhypothesen widerlegt und damit bewiesen, dass die zwei Faktoren einen Einfluss auf die erwartete höchste Klassifikationsstufe einer Institution haben. Von den beiden signifikanten Abhängigkeiten hat der Institutionstyp mit einem Cramérs V-Wert von $\phi_C > 0,3$ einen deutlich höheren–jedoch noch immer moderaten–Einfluss als die Trägerschaft, welche mit $\phi_C > 0,1$ einen vergleichsweise schwachen Einfluss hat.

Institutionstyp Der Einfluss des Institutionstyps besteht darin, wie aus Tabelle 3.3 ersichtlich wird, dass Universitäten viel wahrscheinlicher eine FD-Richtlinie haben als alle anderen Hochschultypen. So haben mehr als die Hälfte aller Universitäten bereits eine dedizierte FD-Richtlinie. Dieser Anteil reduziert sich bei den FHs zu nur einem Viertel. Von den KHs hat sogar nur ca. jede dreizehnte Institution eine FD-Richtlinie. Diese Ungleichverteilung ergibt Sinn, da Universitäten einerseits forschungsorientierter als andere Hochschulen agieren und andererseits zumeist auch ein breiteres fachliches Angebot besitzen: Die forschungsorientiertere Natur von Universitäten hat als logische Konsequenz, dass es wahrscheinlicher ist, dass an Universitäten mehr FD generiert werden und somit damit verbundene externe Faktoren wie z.B. DFG-Förderregularien stärker oder zumindest breitflächiger greifen und damit für die Universität und deren Forschenden von höherer Relevanz sind. Dies würde sich auch mit den Ergebnissen einer früheren Studie aus 2022 decken, welche durch eine Reihe an Interviews aufzeigte, dass Universitätsmitarbeitende eine höhere Dringlichkeit empfinden, auf Auflagen der großen Forschungsförderer zu reagieren, als Mitarbeitende von HAWs [52, S. 9]. Dies wurde unter anderem damit in Verbindung gebracht, dass die HAWs einen Großteil ihrer Drittmittel nicht durch diese Forschungsförderer erhalten, sondern durch Industriekooperationen und andere Geldgeber–welche zumindest zu dem Zeitpunkt der Publikation weniger strikte Auflagen zu FD und FDM besitzt haben [52]. Das breitere fachliche Angebot wiederum erhöht die Wahrscheinlichkeit, dass auch Fachdisziplinen an einer Universität erforscht und unterrichtet werden, die traditionell FD-lastiger sind als andere, was den Bedarf für derartige Richtlinien erhöht. Bei KHs kommt dazu, dass diese Institutionen meistens ein Fächerangebot aufweisen, welches traditionell weniger datenintensive Projekte produziert (z.B. Theologie, Sozialarbeit oder darstellende Kunst), was deren noch niedrigere Rate an FD-Richtlinien und GWP-Verpflichtungen erklären würde.

Trägerschaft Für den zweiten signifikanten Faktor, *Trägerschaft* ist die Erklärung ihrer Wirkung weniger offensichtlich–auch weil diese nur schwacher Natur ist. Es ließe sich die Hypothese aufstellen, dass kirchliche und private Institutionen meist ein spezialisiertes Fachangebot aufweisen und durch nicht-öffentliche (privat-industrielle oder

kirchliche) Drittmittelquellen weniger Drang verspüren, den FD-Auflagen der größeren öffentlichen Forschungsförderorganisationen nachzukommen. Im Umkehrschluss würde die Tatsache, dass die Hauptdrittmittelquellen der öffentlichen Institutionen zumeist das BMBF und andere Förderorganisationen wie die DFG sind [52], bedeuten dass diese einen höheren Druck verspüren, deren striktere FD-Richtlinien zu implementieren.

Für *Bundesland* \times *Richtlinienklassifikation*, hingegen, konnte via Chi-Quadrat-Test mit einem Ergebnis von $p \geq 0,05$ die Nullhypothese bestätigt werden und damit als relevanter Faktor ausgeschlossen werden. Dies ist dadurch erklärbar, dass FD-Richtlinien im Allgemeinen–und im Gegensatz zu Promotions- und Prüfungsordnungen–keine rechtlich notwendige Existenz haben und daher auch weniger von den unterschiedlichen gesetzlichen Anforderungen der Bundesländer beeinflusst werden.

Insignifikante
Faktoren

Was den üblichen Namen für übergreifende Dokumente betrifft, die den Umgang mit FD betreffen, zeigt Abbildung 3.4 auf, dass in Deutschland diesbezüglich bisher wenig Standardisierung vollzogen wurde. Hierbei stimmen die Ergebnisse dieser Arbeit mit denen aus der bereits existierenden Literatur überein [37, 39]. Allerdings ist anzumerken, dass die Verteilung der Dokumentnamen seit 2018, wie sie in Kapitel 2 dargestellt worden sind, noch heterogener geworden ist: Während in 2018 noch 50 % aller Dokumente *Leitfaden* genannt worden sind, sind es in 2024 laut Abbildung 3.4 nur noch 39 %. Die Dokumentnamen *Grundsätze* und *Richtlinie* sanken respektiv von 23 % und 9 % auf jeweils 14 % und 6 %. Im Gegensatz dazu stieg der Anteil an *Policies* von 18 % auf 29 % und wurde durch das neue Auftauchen von *Strategiepapier*, *Leitfaden*, *Empfehlungen* und der Doppelbenennung *Policy & Richtlinie* ergänzt. Es scheint sich hier also auch weiterhin kein Trend zur Standardisierung abzubilden. Dieser Trend sollte weiterhin beobachtet werden. Zukünftige Forschung sollte sich darauf konzentrieren, zu untersuchen, ob die verschiedenen Dokumentbezeichnungen einen Einfluss darauf haben, inwiefern die Institutionen den Auflagen von internationalen Fördervereinen gerecht werden können, wenn diese die Existenz einer FD-Richtlinie bzw. *Policy* voraussetzen. So wurde in früherer Forschung bereits angemerkt, dass die Nutzung anderer Bezeichnungen jenseits von *Richtlinie* und *Policy* zu etwaigen Problemen bei internationalen Projektanträgen führen könnte, da z.B. *Leitlinie* nicht einfach als *Policy* übersetzt werden könnte und damit etwaigen Auflagen nicht gerecht werden würde [37, S. 8].

Name der
Dokumente

Der Chi-Square-Test für die Abhängigkeit zwischen *Dokumentname* und *Dokumentcharakter* ergab, mit einem p-Wert von $p < 0,001$, einen statistisch hochsignifikanten Wert und widerlegte damit die hiermit assoziierte Nullhypothese. Mit einem Cramérs V-Wert $\phi_c > 0,5$ scheint der Name eines Dokumentes einen starken Zusammenhang zu dessen Inhaltscharakter zu haben. Dies ergibt Sinn, da prinzipiell davon ausgegangen werden können sollte, dass z.B. *Empfehlungen* eher einen empfehlenden Charakter hat als *Richtlinie* oder *Policies*. Hier bietet sich dennoch eine lohnenswerte Forschungsmöglichkeit an: Da Dokumente mit teils demselben Namen unterschiedlichen Inhaltscharakter besitzen, wäre es interessant, den Textinhalt aller Dokumente z.B. durch Lemmatisierung zu standardisieren, die Texte zu vektorisieren, die ko-

Inhalt der
Dokumente

sinale Ähnlichkeit der Dokumente zueinander zu berechnen und dann mit einem statistischen Klumpungsverfahren zu kategorisieren. Leider ist der damit verbundene Aufwand in der Vorverarbeitung der Dateien jenseits des Rahmens dieser Arbeit. Wir überlassen dies daher einem zukünftigem Forschungsvorhaben. Abgesehen von der Charakterbewertung wurde der Inhalt der Dokumente mit dem in der Literatur bereits beschriebenen durchschnittlichen Dokumentinhalt verglichen. Hierbei konnten keine erwähnenswerte Unterschiede identifiziert werden. Für eine Inhaltsbeschreibung wird daher, bei Bedarf, auf bereits existierende Literatur verwiesen [37, 39].

3.3.2 Promotionsspezifische Dokumente

Signifikante Faktoren Zwischen den am Kapitelanfang genannten Faktoren und der höchsten Klassifikationsstufe einer Institution für promotionsspezifische Dokumente gab es folgende signifikante Relationen: Mit einem p-Wert von $p < 0,05$ waren die Abhängigkeiten zwischen *Institutionstyp* \times *Spezifische Dokumentklassifikation* und *Bundesland* \times *Spezifische Dokumentklassifikation* statistisch signifikant. Zusätzlich war die Abhängigkeit zwischen *Institutionstyp* und *Richtlinienklassifikation* mit einem p-Wert von $p < 0,001$ statistisch hochsignifikant. Diese zwei Chi-Quadrat-Tests haben dementsprechend die respektiven Nullhypothesen widerlegt und damit bewiesen, dass die zwei Faktoren, bei promotionsspezifischen Dokumenten, einen Einfluss auf die erwartete höchste Klassifikationsstufe einer Institution haben. Von den beiden signifikanten Abhängigkeiten haben beide Faktoren mit einem Cramérs V-Wert von $\phi_c > 0,3$ einen moderaten Einfluss auf die erwartete höchste Klassifikationsstufe einer Institution in Bezug auf promotionsspezifische Dokumente.

Bundesland *Bundesland* hat hierbei einen höheren Einfluss auf das Resultat als der Faktor *Institutionstyp*. Dieser Zusammenhang und die Stärke des Zusammenhangs stehen dabei in starkem Kontrast zu den allgemeinen verwaltungsrechtlichen Dokumenten. Dies lässt sich jedoch dadurch erklären, dass die meisten allgemeinen verwaltungsrechtlichen Dokumente gesetzlich freiwilliger Natur sind. Das bedeutet, dass sie sich weniger stark an den landesspezifischen Gesetzgebungen orientieren müssen. Die Existenz promotionsspezifischer Dokumente (z.B. Promotions- oder Prüfungsordnungen) ist hingegen meist gesetzlich verpflichtender Natur und deren Inhalt daher stark abhängig von den respektiven Bildungsgesetzen der einzelnen Bundesländer. Es würde daher Sinn ergeben, dass *Bundesland* ein starker Indikator für den Inhalt eines solchen Dokumentes ist – auch in Hinblick auf FDM. Dass *Institutionstyp* ebenfalls einen Einfluss auf die Prüfungs- und Promotionsordnungen hat ist ebenfalls wenig verwunderlich. Die Erklärung hierzu ist analog zu der Erklärung für den Einfluss von *Institutionstyp* auf die Klassifikation von allgemeinen FD-Dokumenten. Siehe daher Abschnitt 3.3.1 für Details.

Insignifikante Faktoren Für *Trägerschaft* \times *Richtlinienklassifikation*, hingegen, konnte via Chi-Quadrat-Test mit einem Ergebnis von $p \geq 0,05$ die Nullhypothese bestätigt werden und damit als relevanter Faktor ausgeschlossen werden. Dies ist dadurch erklärbar, dass Promotions- und Prüfungsordnungen – für eine staatliche Anerkennung der Bildungsinstitution – mit

denen aus dem öffentlich-rechtlichen Bereich vergleichbar wenn nicht zu großen Teilen identisch sein müssen. Es gibt daher keinen Grund davon auszugehen, dass sich Institutionen verschiedener Trägerschaften hier unterschiedlich verhalten sollten.

Zusätzlich zu den erwarteten Faktoren wurde durch ein Chi-Quadrat-Test in Abschnitt 3.2.3 auch gezeigt, dass es, mit einem p-Wert von $p < 0,05$, einen statistisch signifikanten Zusammenhang zwischen der höchsten Klassifikationsstufe einer Institution in Bezug auf allgemeine und promotionspezifische Dokumente gibt. Dieser Zusammenhang hat, mit einem Cramérs V-Wert von $\phi_c > 0,3$, eine moderate Stärke. Dies ergibt Sinn, da die Existenz von freiwilligen allgemeinen Richtlinien eine Auswirkung auf den benötigten sowie gewünschten Inhalt einer Promotions- und Prüfungsordnung hat. So zeigt die Existenz einer allgemeinen FD-Richtlinie an, dass eine Institution viel Stellenwert auf richtiges FDM setzt und auch versucht zukunftsorientiert oder zumindest dem wissenschaftlichen Zeitgeist nahestehend zu handeln. Institutionen mit dieser Eigenschaft wären entsprechend wahrscheinlichere Kandidaten dafür, FD-Richtlinien auch in Promotions- und Prüfungsordnungen zu integrieren. Dies würde erklären, warum drei Viertel aller Institutionen mit einer verpflichtenden FD-Richtlinie in einer Promotions- oder Prüfungsordnung auch eine allgemeine Richtlinie zum Umgang mit FD haben.

*Allgemeine
Richtlinien*

DIESE SEITE WURDE ABSICHTLICH LEER GELASSEN

Kapitel 4

Forschungsdaten im Repository der Leibniz Universität Hannover

Dieses Kapitel behandelt die Auswertung von integrierten, begleitenden sowie externen FD in Dissertationen der LUH, die im LUH-Repo veröffentlicht worden sind. Die Arbeit beschränkt sich hierbei exklusiv auf FD, die originelle Primärdaten (PDs) darstellen, die im Rahmen des Promotionsvorhabens entstanden sind. Es wird hierbei überprüft, welcher Anteil an Dissertationen originelle FD beinhaltet, auf welche Art und Weise die FD inkludiert wurden, wie sich diese über die einzelnen Fakultäten verteilen, wie sich diese über die letzten zwölf Jahre entwickelt haben und wie die Existenz von FD in den Metadaten kenntlich gemacht wurden (sowohl im LUH-Repo wie auch in etwaigen externen FD-Repositories).

Thema des Kapitels

Hierfür wird in Abschnitt 4.1 aufgeführt, wie die zu untersuchenden Dissertationen ausgewählt wurden, wie das entsprechende Material gesammelt wurde und mit welchen Methoden dieses daraufhin ausgewertet wurde. In Abschnitt 4.2 werden die entsprechenden Ergebnisse der Materialauswertung dargestellt. Abschließend werden in Abschnitt 4.3 die dargestellten Ergebnisse evaluiert und diskutiert.

Aufbau des Kapitels

4.1 Material & Methoden

In diesem Abschnitt wird das zu untersuchende Material in Abschnitt 4.1.1 und die Methoden der Untersuchung in Abschnitt 4.1.2 dargestellt.

Aufbau des Abschnitts

4.1.1 Material

Als Datengrundlage für dieses Kapitel dient die Metadaten-Datenbank des LUH-Repo, welches durch das Open Archives Initiative Protocol for Metadata Harvesting (OAIPMH) für die Öffentlichkeit frei zugänglich ist [53]. Da sich das Thema dieses Kapitels explizit auf Dissertationen beschränkt, wurden von der Metadaten-Datenbank des LUH-Repas alle Einträge der Sammlung *Dissertationen* die am 21.03.2024 um 10:41 Uhr (UTC+01:00) existierten ($n = 5095$) über die Administrationsübersicht des LUH-Repas heruntergeladen und archiviert [41].

Datengrundlage

Da die zentrale Forschungsfrage dieses Kapitels sich auf den Zeitraum von 2012 bis 2023 beschränkt, wurde diese Liste durch ein Python-Skript [41] auf nur jene Metadateneinträge gefiltert, deren Publikationsjahr in diese Zeitspanne fällt und deren Sperrfrist auch, insofern vorhanden, spätestens 2023 endete ($n = 1898$). Der entsprechende Zeitraum wurde ausgewählt, da 2012 *re3data* gegründet wurde [12] und die darauffolgenden Jahre von vermehrtem Interesse am korrekten Umgang mit FD

Grundmengenbeschreibung

geprägt waren (siehe Kapitel 2 für Details). Die daraus resultierende Dissertationsliste enthält Einträge zu jeder Fakultät der LUH. Aus Gründen der Übersichtlichkeit und des Platzes werden die folgenden Abkürzungen für die Fakultäten der LUH genutzt:

- Fakultät für Architektur und Landschaft (Fakultät ARC)
- Fakultät für Bauingenieurwesen und Geodäsie (Fakultät BAU)
- Fakultät für Elektrotechnik und Informatik (Fakultät INF)
- Juristische Fakultät (Fakultät JUR)
- Fakultät für Maschinenbau (Fakultät MAS)
- Fakultät für Mathematik und Physik (Fakultät MAT)
- Naturwissenschaftliche Fakultät (Fakultät NAT)
- Philosophische Fakultät (Fakultät PHI)
- Wirtschaftswissenschaftliche Fakultät (Fakultät WIWI)

Der Name der Fakultät wird fortan nur noch separat erwähnt, wenn dieser relevant zur Diskussion und Verständlichkeit der Daten erscheint. Der Zeitraum von 2012–2023 wurde für die weitere Bearbeitung wiederum in drei kontinuierliche Zeitintervalle von jeweils vier Jahren aufgeteilt. Die relative sowie die absolute Distribution aller Metadateneinträge nach Zeitraum und Fakultät ist in Tabelle 4.1 gegeben.

Tabelle 4.1: Die Verteilung der Grundmengen-Metadateneinträge nach *Fakultät* \times *Zeitraum* aufgliedert. Absolute Werte in Klammern angegeben.

	2012–2015		2016–2019		2020–2023		Summe	
Fakultät ARC	0,68 %	(13)	1,21 %	(23)	1,37 %	(26)	3,27 %	(62)
Fakultät BAU	1,00 %	(19)	1,79 %	(34)	3,64 %	(69)	6,43 %	(122)
Fakultät INF	1,90 %	(36)	2,63 %	(50)	4,21 %	(80)	8,75 %	(166)
Fakultät JUR	0,00 %	(0)	0,16 %	(3)	0,05 %	(1)	0,21 %	(4)
Fakultät MAS	1,69 %	(32)	1,95 %	(37)	3,48 %	(66)	7,11 %	(135)
Fakultät MAT	5,32 %	(101)	4,48 %	(85)	6,38 %	(121)	16,17 %	(307)
Fakultät NAT	13,22 %	(251)	11,80 %	(224)	15,12 %	(287)	40,15 %	(762)
Fakultät PHI	2,00 %	(38)	1,95 %	(37)	3,37 %	(64)	7,32 %	(139)
Fakultät WIWI	3,16 %	(60)	3,37 %	(64)	4,06 %	(77)	10,59 %	(201)
Summe	28,98 %	(550)	29,35 %	(557)	41,68 %	(791)	100,00 %	(1898)

Für eine Liste aller inkludierter Dissertationsmetadaten, siehe [41].

Stichproben-
ziehung

Diese Liste an Dissertationsmetadaten bildete die Grundmenge für die Ziehung einer mehrschichtigen Zufallsstichprobe. Die Schichten der Zufallsstichprobe entsprachen dabei *Fakultät* \times *Jahresspanne* und ergaben daher insgesamt $9 \times 3 = 27$ Stichprobengruppierungen. Für jede Stichprobengruppierung wurde durch ein Python-Skript [41] eine eigene CSV-Tabellendatei erstellt. Auf die einzelnen Stichprobengruppierungen wurde dann jeweils das Verfahren einer einfachen Stichprobenziehung angewandt.

Bei der Auswahl der Stichproben wurde jeweils ein Konfidenzintervall von 95,00 % und eine Fehlerspanne von 5,00 % zugrunde gelegt. Diese Parameter gewährleisten, dass die Ergebnisse der Stichprobe mit hoher Wahrscheinlichkeit repräsentativ für die gesamte Population sowie der einzelnen Stichprobengruppierungen sind und die Unsicherheit der Schätzungen innerhalb akzeptabler Grenzen bleibt. Um den Prozess der Stichprobenziehung zu automatisieren und eine randomisierte Auswahl zu gewährleisten, wurde eine auf Python basierende Software [42] genutzt, welche im Rahmen dieser Arbeit geschrieben wurde.¹

Die so gezogene Stichprobe besteht aus ca. 76,00 % ($n = 1441$) aller Metadateneinträge der Grundmenge. Die relative sowie die absolute Distribution aller Institutionen in der Stichprobe nach *Fakultät* und *Zeitraum* ist in Tabelle 4.2 gegeben. Der jeweils

Stichproben-
beschreibung

Tabelle 4.2: Die Verteilung der Stichproben-Metadateneinträge nach *Fakultät* \times *Zeitraum* aufgegliedert. Angaben relativ zur Gesamtsumme der Stichprobe. Absolute Werte in Klammern angegeben.

	2012–2015	2016–2019	2020–2023	Summe
Fakultät ARC	0,90 % (13)	1,53 % (22)	1,73 % (25)	4,16 % (60)
Fakultät BAU	1,32 % (19)	2,22 % (32)	4,09 % (59)	7,63 % (110)
Fakultät INF	2,29 % (33)	3,12 % (45)	4,65 % (67)	10,06 % (145)
Fakultät JUR	0,00 % (0)	0,21 % (3)	0,07 % (1)	0,28 % (4)
Fakultät MAS	2,08 % (30)	2,36 % (34)	3,96 % (57)	8,40 % (121)
Fakultät MAT	5,62 % (81)	4,86 % (70)	6,45 % (93)	16,93 % (244)
Fakultät NAT	10,62 % (153)	9,85 % (142)	11,45 % (165)	31,92 % (460)
Fakultät PHI	2,43 % (35)	2,36 % (34)	3,82 % (55)	8,61 % (124)
Fakultät WIWI	3,68 % (53)	3,82 % (55)	4,51 % (65)	12,01 % (173)
Summe	28,94 % (417)	30,33 % (437)	40,74 % (587)	100,00 % (1441)

relative Anteil der Stichprobengruppierungen zu dem entsprechenden Datensatz aus der Grundmenge sowie die Differenz zwischen den respektiven Werten ist, auch nach *Fakultät* und *Zeitraum* aufgegliedert, in Tabelle 4.3 gegeben.

Für die Evaluation, inwiefern die Dissertationen der Stichprobe FD beinhalten oder auf solche verweisen, wurden alle Dateien, die mit Metadateneinträgen assoziiert werden, heruntergeladen. Dieser Prozess wurde dadurch verkompliziert, dass *DSpace 5*, auf welches das LUH-Repo zum Zeitpunkt dieser Arbeit noch basierte, keine eingebaute Möglichkeit bietet, alle Dateien einer Sammlung (jenseits einer schnell erreichten Grenze) oder einer bestimmten Metadatenliste herunterzuladen: weder intern mit administrativen Rechten noch extern durch die Nutzung einer Schnittstelle. Daher wurde im Rahmen dieser Arbeit ein Python-Skript entwickelt, welches alle Links zu den entsprechenden Dateien aus dem öffentlichen Quellcode der Webseiten extrahiert, automatisch herunterlädt und nach dem Metadaten-Handle sortiert [54].

Daten-
sammlung

¹Die Software von KRASSNIG [42] nutzt standardmäßig die Anzahl an Nanosekunden seit dem Beginn der System-Epoche (1970-01-01T00:00:00Z) als Startwert für die Zufallsfunktion. Der genutzte Startwert wird als begleitendes Metadatum der Stichprobe abgespeichert. Die Ziehung ist somit wiederholbar und das Datum der Ziehung verifizierbar.

Tabelle 4.3: Die Verteilung der Stichproben-Metadateneinträge nach *Fakultät* × *Zeitraum* aufgliedert. Angaben relativ zu dem jeweiligen *Fakultät* × *Zeitraum*-Wert der Grundmenge. Absolute Differenzwerte in Klammern angegeben.

	2012–2015		2016–2019		2020–2023		Alle	
Fakultät ARC	100,00 %	(0)	95,65 %	(-1)	96,15 %	(-1)	96,77 %	(-2)
Fakultät BAU	100,00 %	(0)	94,12 %	(-2)	85,51 %	(-10)	90,16 %	(-12)
Fakultät INF	91,67 %	(-3)	90,00 %	(-5)	83,75 %	(-13)	87,35 %	(-21)
Fakultät JUR	—	(0)	100,00 %	(0)	100,00 %	(0)	100,00 %	(0)
Fakultät MAS	93,75 %	(-2)	91,89 %	(-3)	86,36 %	(-9)	89,63 %	(-14)
Fakultät MAT	80,20 %	(-20)	82,35 %	(-15)	76,86 %	(-28)	79,48 %	(-63)
Fakultät NAT	60,96 %	(-98)	63,39 %	(-82)	57,49 %	(-122)	60,37 %	(-302)
Fakultät PHI	92,11 %	(-3)	91,89 %	(3)	85,94 %	(-9)	89,21 %	(-15)
Fakultät WIWI	88,33 %	(-7)	85,94 %	(-9)	84,42 %	(-12)	86,07 %	(-28)
Alle	75,82 %	(-133)	78,46 %	(-120)	74,21 %	(587)	75,92 %	(-457)

Hierbei wurden insgesamt 1 480 Dateien zur Weiterverarbeitung gefunden, heruntergeladen und sortiert.

Zahlenspiegel
der LUH

Zusätzlich zu den oben aufgeführten Dissertationen wurden auch die von der LUH veröffentlichten Zahlenspiegel für den zu untersuchenden Zeitraum gesammelt, da diese die jährliche Gesamtzahl veröffentlichter Dissertationen enthalten. Hierbei wurden die Zahlenspiegel von 2013–2023 ausgesucht, da diese jeweils das vorherige Jahr betreffen [55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65]. Die aktuellen Daten für 2023, bzw. der Zahlenspiegel aus 2024 wurde zum Zeitpunkt dieser Arbeit noch nicht veröffentlicht.

4.1.2 Methoden

Klassifikation

Die in Abschnitt 4.1.1 gesammelten Dissertationsdateien wurden dann, um die zentrale Forschungsfrage dieses Kapitels zu beantworten, danach klassifiziert, ob und auf welche Art und Weise sie primäre FD beinhalten: FD konnten entweder in die PDF-Datei integriert, als Begleitdaten im LUH-Repo eingereicht oder auf ein externes FD-Repositorium hochgeladen worden sein [3]. Damit diese Klassifikation stattfinden konnte, musste jedoch zuerst bestimmt werden, welcher Inhalt als FD gewertet wird (hierbei orientierte sich diese Arbeit an [14, 66]) und wo sich diese typischerweise im Dokument befinden. Hierfür wurden von jeder Fakultät, gleichmäßig auf die drei Zeiträume aufgeteilt, zwölf zufällige Dissertationen ausgewählt und vorläufig evaluiert. Bei Stichprobengruppierungen von weniger als vier Dissertationen wurden stattdessen alle Dokumente vorläufig ausgewertet.

Klassifikations-
hierarchie

Bei dieser Auswertung wurde ein provisorisches Klassifikationssystem aufgebaut, welches für den Rest der Arbeit beibehalten wurde. Hierbei wurden FD in drei hierarchische Stufen eingeteilt. Diese reichen von *Stufe 1*, welche eindeutige und zweifelsfreie PDs beinhalten bis zu *Stufe 3*, welche Daten beinhalten, die entweder kom-

promittiert worden sind (z.B. durch starke Kompression), keine besondere Leistung darstellen (z.B. Fragebögen ohne Antworten) oder durch diese Abschlussarbeit kaum auf Originalität überprüfbar waren (z.B. Gensequenzen, da diese häufig scheinbar nachgenutzt wurden). Unter *Stufe 2* befinden sich jene FD, welche zwar originell sind, jedoch weniger direkte Wiederverwendbarkeit oder Qualität im Vergleich zu FD aus *Stufe 1* haben. Es folgt eine Auflistung der verschiedenen FD-Klassifikationen der entsprechenden Klassifikationsstufen.

Stufe 1: rohe Beobachtungs- / Messdaten, unkomprimierte Rohbilder, Videos, Skripte / Software, Transkriptionen von Interviews, Beantwortungen von Fragebögen

Stufe 2: Pseudocode, Algorithmen, komprimierte Bilder von Gelfärbungen²

Stufe 3: komprimierte Fotografien und Scans, Spektraldiagramme, Gensequenzen, unbeantwortete Fragebögen, Leitfäden, Montagezeichnungen

Bei der vorläufigen Testklassifikation konnten keine bestimmten Teile eines Dokumentes vollständig ausgeschlossen werden. Während die meisten hochstufigen FD in dem jeweiligen Appendix zu finden waren, waren z.B. FD der zweiten und dritten Stufe zu großen Teilen im gesamten Dokument verteilt. Auch FD aus *Stufe 1* waren teilweise in anderen Bereichen der Dissertationen zu finden. Dies galt auch für externe Forschungsdaten. So wurde teilweise in der Präambel darauf hingewiesen, dass einige oder alle FD auf ein externes Repositorium hochgeladen wurden. Teilweise wurden diese externen Datensätze aber auch erst an der jeweils relevanten Stelle im Hauptteil des Dokumentes zitiert. Für den restlichen Verlauf der Klassifikation wurde daher beschlossen, dass sämtliche Seiten der Dissertationen zumindest kurzzeitig begutachtet werden müssen.

Ort der FD

Nach Abschluss der vorläufigen Testklassifikation und Aufbau des hierarchischen Klassifikationssystems wurde dann wie folgt vorgegangen: (i) Es wurden alle PDF-Dateien einzeln evaluiert. (ii) Insofern ein Typ an FD im Dokument gefunden wurde, so wurde der FD-Typ und die Publikationsart der FD in der zur Stichprobenschicht zugehörigen CSV-Tabellendatei vermerkt. (iii) Hierbei wurde für interne und beigefügte FD nur der Typ und nicht die Seite innerhalb des Dokumentes vermerkt. Für extern publizierte FD wurden zusätzlich noch die jeweilig dazugehörigen Seiten und die Art des externen Repositoriums vermerkt (z.B. Git-Repositorium oder dediziertes FD-Repositorium). (iv) Bei externen FD wurden zusätzlich die DOI, das relevante Stichwort oder die dazugehörige Domäne notiert und in einer separaten Datei eingetragen. (v) Diese Liste wurde dann am Ende der Evaluation genutzt, um ein Bash-Skript [41] zu erstellen, welches automatisch den Text aller Dateien nach den notierten Wörtern durchsucht, um etwaige übersehene externe FD im Nachhinein noch erfassen zu können. (vi) Zusätzlich wurde für jede Dissertation auch eingetragen, ob für sie überhaupt PDs produziert worden sind; diese Information wurden wiederum genutzt, um relative Werte zu der jeweiligen Gesamtsumme aller Dissertationen minus jenen ohne produzierte PDs zu erstellen.

Klassifikationsstrategie

²Für komprimierte Bilder von Gelfärbungen wurde auf Anraten einer Wissenschaftlerin aus dem Bereich *Life Science* eine Ausnahme gemacht und als FD der zweiten Stufe kategorisiert [67].

- Gesamt-
klassifikation* Nach Beendigung der Klassifikationsarbeit wurden jeder Dissertation jeweils vier Werte zugeordnet:
- die höchste Klassifikationsstufe aller gefundenen FD der Dissertation
 - die höchste Klassifikationsstufe aller gefundenen FD die in der PDF-Datei der Dissertation integriert waren
 - die höchste Klassifikationsstufe aller gefundenen FD die der Dissertation als separate Datei beigelegt wurden
 - die höchste Klassifikationsstufe aller gefundenen FD die auf einer externen Plattform hochgeladen wurden.
- Statistische
Auswertung* Nach der Klassifizierung aller Dateien wurden die vollständig evaluierten CSV-Dateien der Stichprobengruppierungen durch ein Python-Skript [41] einerseits in kombinierte Fakultätstabellen und andererseits in eine fakultätsübergreifende Gesamttabelle zusammengeführt. Die vorhandenen Metadaten wurden dann auf plausible Faktoren untersucht, die einen etwaigen Einfluss auf die Präsenz, Art und Publikationsform von FD haben. Für die Kreuzprodukte aller vermuteter Faktoren sowie für die Ergebnisse der Klassifikationsarbeit aller FD-Publikationsarten wurden Chi-Quadrat-Tests für Unabhängigkeit durchgeführt, um zu überprüfen, ob statistisch signifikante Relationen zwischen den jeweiligen Faktoren bzw. Ergebnissen besteht [45, 46]. Hierbei wurden für alle zu überprüfenden Relationen die Nullhypothese angenommen: I.e. für die Kombination *Faktor A* \times *Faktor B* wird angenommen, dass *Faktor A* keinen Einfluss auf *Faktor B* hat und dass, bedingt durch die symmetrische Natur des Chi-Quadrat-Tests, auch andersherum kein Einfluss stattfindet. Siehe Abschnitt 3.1.2 für Details zu der Auswertung von Chi-Quadrat-Tests. Da p-Werte nichts über die Stärke einer Abhängigkeit aussagen, wurden für alle Testergebnisse mit $p < 0,05$ zusätzlich noch der respektive Cramérs V-Wert (ϕ_c) berechnet, um zu überprüfen, wie stark die statistisch signifikante Abhängigkeit ist [47]. Siehe Abschnitt 3.1.2 für Details zu der Auswertung von Cramérs V-Werten.
- Zahlenspiegel
der LUH* Die in Abschnitt 4.1.1 gesammelten Zahlenspiegel der LUH wurden dann wie folgt ausgewertet: (i) Es wurden alle Gesamtanzahlen der Dissertationen pro Fakultät für jeden Jahrgang extrahiert und in eine gemeinsame Tabelle eingetragen, wo sie in die Jahresgruppen 2012–2015, 2016–2019 und 2020–2023 geklumpt wurden. (ii) Da die offiziellen Zahlen des Jahres 2023 zum Zeitpunkt dieser Arbeit noch nicht veröffentlicht waren, wurden diese durch das arithmetische Mittel der Fakultätswerte aus den Jahren 2012–2022 simuliert, um vergleichbare Jahreswerte für jede Fakultät zu derivieren. (iii) Wäre es durch diese Methode zu einem Wert kommen, der kleiner gewesen ist, als der entsprechende Wert aus dem LUH-Repo, so wäre in diesem Fall der Wert aus dem LUH-Repo übernommen worden. (iv) Wenn Werte, zumindest in Teilen, auf dieser Simulation basieren, so wird dies mit einem Sternchen (*) markiert. Die resultierende Verteilung ist in Tabelle 4.4 zu sehen. Die *Fakultät* \times *Jahresgruppe*-Zahlen wurden dann mit den Daten aus dem LUH-Repo verglichen. Aus diesem Vergleich wurde dann jeweils abgeleitet, zu welchem Anteil die Promovierenden der verschiedenen Fakultäten das LUH-Repo zur Publikation ihrer Dissertationen nutzen.

Tabelle 4.4: Die Verteilung der Dissertationen laut den Zahlenspiegeln der LUH nach *Fakultät* \times *Zeitraum* aufgegliedert. Angaben relativ zu der Gesamtsumme der Zahlenspiegel-Dissertationen. Absolute Werte in Klammern angegeben. Spalten, die zumindest teilweise auf simulierten Werten basieren, sind mit einem Sternchen (*) markiert.

	2012–2015	2016–2019	2020–2023*	Summe*
Fakultät ARC	0,95 % (40)	0,81 % (34)	0,76 % (32)	2,53 % (106)
Fakultät BAU	2,00 % (84)	2,88 % (121)	2,79 % (117)	7,67 % (322)
Fakultät INF	3,15 % (132)	3,31 % (139)	3,46 % (145)	9,91 % (416)
Fakultät JUR	2,41 % (101)	1,53 % (64)	2,03 % (85)	5,96 % (250)
Fakultät MAS	5,24 % (220)	6,29 % (264)	6,79 % (285)	18,33 % (769)
Fakultät MAT	4,58 % (194)	3,72 % (156)	3,62 % (152)	11,92 % (500)
Fakultät NAT	9,10 % (384)	8,77 % (368)	7,96 % (334)	25,83 % (1084)
Fakultät PHI	3,36 % (142)	3,77 % (158)	3,19 % (134)	10,32 % (433)
Fakultät WIWI	2,67 % (107)	2,76 % (116)	2,10 % (88)	7,53 % (316)
Summe	33,46 % (1404)	33,84 % (1420)	32,70 % (1372)	100,00 % (4196)

Alle statistischen Auswertungen wurden via *R* ausgeführt [41, 48, 49]. Für die Klassifikationsarbeit sowie für alle anderweitigen manuellen Bearbeitungen der Datentabellen wurde *LibreOffice Calc* genutzt [50].

Software

4.2 Resultate

In diesem Abschnitt werden die Resultate der Zahlenspiegelauswertung und der LUH-Repo Datenklassifizierung sowie deren statistische Auswertung dargestellt. In Abschnitt 4.2.1 werden die Zahlen der LUH-Zahlenspiegel ausgewertet und in Relation zu den Zahlen des LUH-Repos gesetzt. Hiermit wird eine Nutzungsrate für die Fakultäten über die einzelnen Zeitgruppen hergestellt. Fakultäten, deren Verhalten sich ähneln, werden in entsprechende Gruppen zusammengefasst. In Abschnitt 4.2.2 werden mögliche Faktoren für die Existenz und Publikationsart von FD in Dissertationen identifiziert und auf Unabhängigkeit voneinander überprüft. In Abschnitt 4.2.3 wird die manuelle Klassifikation der Dissertationen genutzt, um zu beurteilen, wie viele der Dissertationen überhaupt PDs produziert haben. Dieser Wert wird für die Auswertung des FD-Verhaltens in den darauffolgenden Abschnitten genutzt. In Abschnitt 4.2.4 werden die FD-Klassifikationsresultate in einen zeitlichen Kontext dargestellt und gezeigt, inwiefern sich diese über die verschiedenen Zeitgruppen verhalten. In Abschnitt 4.2.5 wird der etwaige Faktor *Sprache* untersucht und in Relation zu den Fakultäten sowie den einzelnen Zeitgruppen gebracht. In Abschnitt 4.2.6 werden die Resultate der Klassifikationsarbeit im Kontext der Fakultäten dargestellt. Hier wird untersucht, ob und, falls ja, inwiefern sich die Fakultäten in Bezug auf FD-Verhalten unterscheiden. In Abschnitt 4.2.7 werden die Ergebnisse zu den Metadaten gezeigt, die bei den gefundenen externen FD gefunden worden sind.

Aufbau des Abschnitts

4.2.1 Nutzungsrate des LUH-Repos

Zahlenspiegel

Die Auswertung des Zahlenspiegels und der Grundmenge dieser Arbeit ergab, dass nur 45,23 % ($n = 1\,898, \Delta = -2\,298$) aller Dissertationen im Zeitraum von 2012–2023 im LUH-Repo erst- oder zweitveröffentlicht worden sind. Der relative Anteil und die absolute Differenz zwischen der Grundmenge und den Werten aus den Zahlenspiegeln der LUH, nach *Zeitgruppe* und *Fakultät* aufgegliedert, sind in Tabelle 4.6 gegeben.

Tabelle 4.5: Der Anteil der Grundmenge nach *Fakultät* × *Zeitraum* aufgegliedert relativ zu der respektiven *Fakultät* × *Zeitgruppe*-Gesamtanzahl aller publizierten Dissertationen. Absolute Differenzwerte in Klammern angegeben. Spalten, die zumindest teilweise auf simulierten Werten basieren, sind mit einem Sternchen (*) markiert.

	2012–2015	2016–2019	2020–2023*	Summe*
Fakultät ARC	32,50 % (-27)	67,65 % (-11)	81,25 % (-6)	58,49 % (-44)
Fakultät BAU	22,62 % (-65)	28,10 % (-87)	58,97 % (-48)	37,89 % (-200)
Fakultät INF	27,27 % (-96)	35,97 % (-89)	55,17 % (-65)	39,90 % (-250)
Fakultät JUR	0,00 % (-101)	4,69 % (-61)	1,18 % (-84)	1,60 % (-246)
Fakultät MAS	14,55 % (-188)	14,02 % (-227)	23,16 % (-219)	17,56 % (-634)
Fakultät MAT	52,06 % (-93)	54,49 % (-71)	79,61 % (-31)	61,40 % (-193)
Fakultät NAT	65,36 % (-133)	60,87 % (-144)	85,93 % (-47)	70,30 % (-322)
Fakultät PHI	26,76 % (-104)	23,42 % (-121)	47,76 % (-70)	32,10 % (-294)
Fakultät WIWI	56,07 % (-47)	55,17 % (-52)	87,50 % (-11)	63,61 % (-115)
Summe	39,17 % (-854)	39,23 % (-863)	57,65 % (-581)	45,23 % (-2298)

Mit einem arithmetische Mittel aller *Fakultät* × *Zeitgruppe*-Kombinationen von $\bar{x} = 42,99\%$ und einer Standardabweichung von $s = 26,15\%P$ ist die relative Nutzung des LUH-Repos für die Veröffentlichung von Dissertationen unter den Fakultäten und Jahresgruppen sehr ungleich.

Nutzer-
gruppierung

Die Fakultäten lassen sich hierbei durch ihre durchschnittliche Nutzungsrate in drei relativ klar abgegrenzte Gruppen einteilen: Geringnutzer ($\bar{x} < 20,00\%$), Intermediärnutzer ($\bar{x} \approx 33,00\%$) und Intensivnutzer ($\bar{x} \approx 66,00\%$). Die Geringnutzer bestehen aus Fakultät JUR ($\bar{x} = 1,95\%, s = 2,44\%P$) und Fakultät MAS ($\bar{x} = 17,24\%, s = 5,13\%P$). Die Intermediärnutzer bestehen aus Fakultät BAU ($\bar{x} = 36,56\%, s = 19,60\%P$), Fakultät INF ($\bar{x} = 39,47\%, s = 14,28\%P$) und Fakultät PHI ($\bar{x} = 32,71\%, s = 13,15\%P$). Die Intensivnutzer bestehen aus Fakultät ARC ($\bar{x} = 60,47\%, s = 25,16\%P$), Fakultät MAT ($\bar{x} = 62,23\%, s = 15,07\%P$), Fakultät NAT ($\bar{x} = 70,83\%, s = 13,29\%P$) und Fakultät WIWI ($\bar{x} = 65,41\%, s = 19,14\%P$).

Zeitlicher
Trend

Die durchschnittliche Nutzungsrate durch die Fakultäten hat sich 2012–2023* um 86,86 % ($s = 47,06\%P$) erhöht. So lag 2012–2015 für 66,67 % ($n = 6$) der Fakultäten die Nutzungsrate noch unter 50,00 %, während dies in 2020–2023* nur noch für 33,33 % ($n = 3$) der Fall. Hierbei ist zusätzlich noch anzumerken, dass eine dieser Fakultäten–nämlich Fakultät PHI–nur 2,24 %P von 50,00 % entfernt war.

4.2.2 Mögliche relevante Faktoren

Die Untersuchung der Metadaten-Datenbank auf mögliche Faktoren, die Einfluss auf die Existenz von FD haben könnten ergab, zusätzlich zu *Zeitgruppe* und *Fakultät*, nur noch *Sprache*.

Faktoren

Zur Überprüfung, ob die zu untersuchenden Faktoren voneinander abhängig sind, wurden für die Kreuzprodukte aller Faktorenkombinationen Chi-Quadrat Tests der Unabhängigkeit durchgeführt. Hierbei zeigte sich, dass *Zeitgruppe* \times *Fakultät* ($\chi^2(16, n = 1441) = 30,12, p = 0,02, \phi_C = 0,10$), *Zeitgruppe* \times *Sprache* ($\chi^2(6, n = 1441) = 81,20, p = 2,01 \cdot 10^{-15}, \phi_C = 0,17$) und *Fakultät* \times *Sprache* ($\chi^2(24, n = 1441) = 239,31, p = 2,15 \cdot 10^{-37}, \phi_C = 0,24$) alle statistisch signifikant voneinander abhängig sind. Die Effektstärken sind dabei leicht unterschiedlich ausgeprägt aber schwacher Natur.

Unabhängigkeit der Faktoren

4.2.3 Rate an erzeugten Primärdaten

Die Evaluation aller Stichproben-Einträge ergab, dass nur 86,81 % ($n = 1252$) der Stichproben-Dissertationen Dissertationen mit produzierten Primärdaten (mit oder ohne Publikation dieser Primärdaten) (PD-Dissertationen) sind. Die relative und absolute Verteilung nach *Zeitgruppe* und *Fakultät* ist hierfür in Tabelle 4.6 gegeben. Mit $\bar{x} = 77,98\%$, $s = 28,49\%$ P für die gesamte *Fakultät* \times *Zeitgruppe*-Verteilung

PD-Produktion

Tabelle 4.6: Anteil an PD-Dissertationen aus der Stichprobe. Angaben relativ zu der respektiven *Fakultät* \times *Zeitgruppe*-Gesamtanzahl. Absolute Werte in Klammern angegeben.

	2012–2015		2016–2019		2020–2023		Alle	
Fakultät ARC	92,31 %	(12)	81,82 %	(18)	84,00 %	(21)	85,00 %	(51)
Fakultät BAU	94,74 %	(18)	90,63 %	(29)	98,31 %	(58)	95,45 %	(105)
Fakultät INF	96,97 %	(32)	91,11 %	(41)	100,00 %	(67)	96,55 %	(140)
Fakultät JUR	—	(0)	0,00 %	(0)	0,00 %	(0)	0,00 %	(0)
Fakultät MAS	96,67 %	(29)	100,00 %	(34)	100,00 %	(57)	99,17 %	(120)
Fakultät MAT	79,01 %	(64)	81,43 %	(57)	100,00 %	(93)	87,70 %	(214)
Fakultät NAT	100,00 %	(153)	98,59 %	(140)	98,18 %	(162)	98,91 %	(455)
Fakultät PHI	68,57 %	(24)	58,82 %	(20)	49,09 %	(27)	57,26 %	(71)
Fakultät WIWI	52,83 %	(28)	63,64 %	(35)	50,77 %	(33)	55,49 %	(96)
Alle	86,33 %	(360)	85,58 %	(374)	88,25 %	(517)	86,88 %	(1252)

ist die Distribution von PD-Dissertationen im LUH-Repo unter den Fakultäten und Jahresgruppen relativ ungleich.

Die Fakultäten lassen sich auch hier durch ihre durchschnittliche Anzahl an PD-Dissertationen in drei relativ klar abgegrenzte Gruppen einteilen: Nicht-Erzeuger ($\bar{x} \approx 0,00\%$), Intermediärerzeuger ($\bar{x} \approx 50,00\%$) und Intensiverzeuger ($\bar{x} \approx 90,00\%$). Die Nicht-Erzeuger bestehen nur aus Fakultät JUR ($\bar{x} = 0,00\%$, $s = 0,00\%$ P). Die Intermediärerzeuger bestehen aus Fakultät PHI ($\bar{x} = 58,83\%$, $s = 9,74\%$ P) und Fakultät WIWI ($\bar{x} = 55,75\%$, $s = 6,91\%$ P). Die Intensiverzeuger bestehen aus Fakultät ARC

Gruppierung

($\bar{x} = 86,04\%$, $s = 5,53\%P$), Fakultät BAU ($\bar{x} = 94,56\%$, $s = 3,84\%P$), Fakultät INF ($\bar{x} = 96,03\%$, $s = 4,52\%P$), Fakultät MAS ($\bar{x} = 98,89\%$, $s = 1,92\%P$), Fakultät MAT ($\bar{x} = 86,81\%$, $s = 11,48\%P$) und Fakultät NAT ($\bar{x} = 98,92\%$, $s = 0,95\%P$).

4.2.4 Zeitliche Entwicklung der Forschungsdaten

Klassifikation
Zeitgruppen

Von den 1252 PD-Dissertationen haben 61,34% ($n = 768$) zumindest Teile dieser PDs veröffentlicht. Bei einer Beschränkung auf FD von *Stufe 1* und *Stufe 2*, haben nur 31,55% ($n = 395$) zumindest Teile ihrer PDs veröffentlicht. Von allen PD-Dissertationen waren insgesamt nur 20,13% ($n = 252$) als *Stufe 1* klassifiziert. Die relative und absolute Verteilung der Klassifikationsstufen ist für *FD-Publikationsart* \times *Zeitgruppe* \times *Klassifikationsstufe* von PD-Dissertationen in Tabelle 4.7 gegeben. Dieselbe

Tabelle 4.7: FD-Klassifizierung der PD-Dissertationen aus der Stichprobe nach *Publikationsart* \times *Klassifikationsstufe* \times *Jahresgruppe* aufgegliedert. Angaben relativ zu der Gesamtanzahl der jeweiligen Zeitgruppe. Absolute Werte in Klammern angeben.

		2012-2015		2016-2019		2020-2023		Alle	
Intern	Stufe 1	13,89%	(50)	14,44%	(54)	10,62%	(55)	12,70%	(159)
	Stufe 2	13,33%	(48)	14,44%	(54)	13,13%	(68)	13,58%	(170)
	Stufe 3	36,94%	(133)	33,96%	(127)	24,13%	(125)	30,75%	(385)
	Keine	35,83%	(129)	37,17%	(139)	52,12%	(270)	42,97%	(538)
Beilage	Stufe 1	1,39%	(5)	2,67%	(10)	1,35%	(7)	1,76%	(22)
	Stufe 2	0,00%	(0)	0,00%	(0)	0,00%	(0)	0,00%	(0)
	Stufe 3	0,56%	(2)	0,27%	(1)	0,58%	(3)	0,48%	(6)
	Keine	98,06%	(353)	97,06%	(363)	98,07%	(508)	97,76%	(1224)
Extern	Stufe 1	1,11%	(4)	3,21%	(12)	14,29%	(74)	7,19%	(90)
	Stufe 2	0,00%	(0)	0,00%	(0)	0,19%	(1)	0,08%	(1)
	Stufe 3	0,00%	(0)	0,27%	(1)	0,19%	(1)	0,16%	(2)
	Keine	98,89%	(356)	96,52%	(361)	85,33%	(442)	92,57%	(1159)
Alle	Stufe 1	15,28%	(55)	19,79%	(74)	23,75%	(123)	20,13%	(252)
	Stufe 2	13,33%	(48)	12,30%	(46)	9,46%	(49)	11,42%	(143)
	Stufe 3	36,67%	(132)	33,42%	(125)	22,39%	(116)	29,79%	(373)
	Keine	34,72%	(125)	34,49%	(129)	44,40%	(230)	38,66%	(484)

Verteilung für alle Dissertationen ist in Tabelle B.3 gegeben.

Allgemein

Für PD-Dissertationen sind die respektiven Anteile pro Zeitgruppe für *Stufe 1* ($\bar{x} = 19,60\%$, $s = 4,24\%P$), *Stufe 2* ($\bar{x} = 11,70\%$, $s = 2,01\%P$), *Stufe 3* ($\bar{x} = 30,83\%$, $s = 7,48\%P$) und *Keine* ($\bar{x} = 37,87\%$, $s = 5,66\%P$) alle sehr konsistent in ihren Ausmaßen und haben sich über die Jahre nur wenig verändert: Der Anteil von *Keine FD* und *Stufe 1* stieg über die Zeitgruppen hinweg fast stetig an, während der Anteil an *Stufe 2* und *Stufe 3* abnahm. Diese Interaktion ist mit $\chi^2(6, n = 1252) = 35,17$, $p = 3,99 \cdot 10^{-6} < 0,001$, $\phi_C = 0,12$ hochsignifikant aber schwach im Effekt.

Für integriert publizierte FD sind die respektiven Anteile pro Zeitgruppe für *Stufe 1* ($\bar{x} = 12,98\%$, $s = 2,07\%P$), *Stufe 2* ($\bar{x} = 13,63\%$, $s = 0,71\%P$), *Stufe 3* ($\bar{x} = 31,68\%$, $s = 6,70\%P$) und *Keine* ($\bar{x} = 41,71\%$, $s = 9,05\%P$): An den entsprechenden Standardabweichungen lässt sich erkennen, dass sich die Anteile über die Zeitgruppen hinweg geändert haben. Die Anteile von *Keine FD*, *Stufe 1* und *Stufe 3* waren über die ersten zwei Zeitgruppen stabil, stiegen dann aber für *Keine FD* stark an, während *Stufe 1* und *Stufe 3* stark abnahmen. *Stufe 2* verblieb über alle Jahresgruppen hinweg stabil. Diese Interaktion ist mit $\chi^2(6, n = 1252) = 33,78$, $p = 7,42 \cdot 10^{-6} < 0,001$, $\phi_C = 0,12$ hochsignifikant aber schwach im Effekt.

Integrierte FD

Für begleitende FD verändern sich die respektiven Anteile pro Zeitgruppe für *Stufe 1* ($\bar{x} = 1,80\%$, $s = 0,75\%P$), *Stufe 2* ($\bar{x} = 0,00\%$, $s = 0,00\%P$), *Stufe 3* ($\bar{x} = 0,47\%$, $s = 0,17\%P$) und *Keine* ($\bar{x} = 97,73\%$, $s = 0,58\%P$) fast überhaupt nicht. Entsprechend ist diese Interaktion mit $\chi^2(4, n = 1252) = 3,08$, $p = 0,54 > 0,05$ auch nicht signifikant.

Begleitende FD

Für externe FD sind die respektiven Anteile pro Zeitgruppe für *Stufe 2* und *Stufe 3* mit Maximalwerten von unter $0,60\%$ und $s < 0,15\%P$ vernachlässigbar. Für *Stufe 1* ($\bar{x} = 6,20\%$, $s = 7,08\%P$) und *Keine* ($\bar{x} = 0,94\%$, $s = 7,24\%P$) zeichnet sich jedoch auch eine eindeutige Veränderung über die Zeitgruppen hinweg ab: Während *Keine FD* über die Zeitgruppen hinweg um $13,56\%P$ abnimmt, steigt *Stufe 1* um $13,18\%P$ auf einen Maximalwert von $14,29\%$ an. Die verbleibenden $0,38\%P$ teilten sich dabei gleichmäßig auf *Stufe 2* und *Stufe 3* mit jeweils einer PD-Dissertationen auf. Diese Interaktion ist mit $\chi^2(6, n = 1252) = 70,45$, $p = 3,30 \cdot 10^{-13} < 0,001$, $\phi_C = 0,17$ hochsignifikant aber schwach im Effekt. Hier ist allerdings anzumerken, dass sowohl der Signifikanzwert wie auch die Einflusstärke dieser Interaktion höher bzw. stärker ist als die Werte der anderen bisher untersuchten Interaktionen.

Externe FD

Zusammenfassend lassen sich hier folgende Punkte festhalten: Der Anteil publizierter FD der *Stufe 1* hat insgesamt über die Zeitgruppen zugenommen. Dafür wurden die FD aber weniger integriert und vermehrt extern publiziert. Die Option FD als begleitende Dateien zu publizieren, wurde von einem sehr kleinen Nutzerkreis konsistent genutzt.

Zusammenfassung

4.2.5 Sprache und Forschungsdaten

Von allen Dissertationen wurden $54,55\%$ ($n = 786$) auf Deutsch und $45,11\%$ ($n = 650$) auf Englisch verfasst. Die restlichen $0,35\%$ ($n = 5$) wurden entweder auf Spanisch oder in mehreren Sprachen verfasst. Für PD-Dissertationen ändern sich diese Anteile auf jeweils $55,99\%$ ($n = 701$), $43,69\%$ ($n = 547$) und $0,32\%$ ($n = 4$). Für die relative und absolute Verteilung der PD-Dissertationen nach *Sprache* \times *Zeitgruppe*, siehe Tabelle 4.8. Dieselbe Aufteilung für alle Dissertationen ist in Tabelle B.1 gegeben.

Sprache

Für PD-Dissertationen sind die respektiven Anteile pro Zeitgruppe für *Deutsch* ($\bar{x} = 58,02\%$, $s = 14,71\%P$), *Englisch* ($\bar{x} = 41,72\%$, $s = 14,28\%P$) und *Andere* ($\bar{x} = 0,19\%$, $s = 0,33\%P$). Die hohen Standardabweichungen für *Englisch* und *Deutsch* sind dem verschuldet, dass für die ersten zwei Zeitgruppen *Deutsch* vergleichsweise konsistent mit $\bar{x} = 66,28\%$ und *Englisch* mit $\bar{x} = 33,72\%$ vertreten war (jeweils $s =$

Zeitlicher Trend

Tabelle 4.8: FD-Klassifizierung der PD-Dissertationen aus der Stichprobe nach *Sprache* × *Zeitgruppe* aufgegliedert. Angaben relativ zu der Gesamtanzahl der jeweiligen Zeitgruppe. Absolute Werte in Klammern angegeben.

	2012-2015	2016-2019	2020-2023	Alle
Deutsch	69,72 % (251)	62,83 % (235)	41,51 % (215)	55,99 % (701)
Englisch	30,28 % (109)	37,17 % (139)	57,72 % (299)	43,69 % (547)
Andere	0,00 % (0)	0,00 % (0)	0,77 % (4)	0,32 % (4)

4,87%P), in der dritten Zeitgruppe dann aber ein Wechsel der Mehrheit von *Deutsch* (41,51%) auf *Englisch* (57,72%) stattgefunden hat. Die Interaktion zwischen *Sprache* und *Zeitgruppe* ist mit $\chi^2(6, n = 1252) = 81,20, p = 2,01 \cdot 10^{-15} < 0,01, \phi_C = 0,17$ statistisch hochsignifikant aber schwach im Effekt.

Fakultäten Auch die Interaktion zwischen *Sprache* und *Fakultät* ist mit $\chi^2(21, n = 1252) = 201,42, p = 1,90 \cdot 10^{-31} < 0,01, \phi_C = 0,23$ statistisch hochsignifikant aber schwach im Effekt. Diese Interaktion ist durch die stark unterschiedliche Verteilung von Sprachwahl unter den Fakultäten bedingt. So hat *Deutsch* innerhalb der Fakultäten einen Anteil von $\bar{x} = 59,02\%, s = 23,48\%$ und *Englisch* von $\bar{x} = 40,28\%, s = 24,34\%$ P mit respektiven Maximalwerten von 91,55% und 71,88%. Zusätzlich dazu ändert sich die Sprachverteilung innerhalb der Fakultäten je nach *Zeitgruppe* unterschiedlich stark, wie in Abbildung 4.1 visualisiert wird.

Allgemeine FD Für die Beziehung zwischen *Sprache* und *Allgemeine FD* ist anzumerken, dass 67,56% aller *Stufe 3* PD-Dissertationen auf Deutsch und 32,17% auf Englisch geschrieben wurden. Für *Stufe 1* sind diese Anteile beide jeweils 49,60%. Die Abhängigkeit zwischen *Sprache* und *Allgemeine FD* ist mit $\chi^2(9, n = 1252) = 36,32, p = 3,48 \cdot 10^{-5} < 0,01, \phi_C = 0,10$ statistisch hochsignifikant aber sehr schwach im Effekt.

Externe FD Für *Externe FD* ist diese Verteilung umgekehrt und stärker ausgeprägt: Für PD-Dissertationen mit externen FD die mit *Stufe 1* klassifiziert wurden, wurden nur 25,56% auf Deutsch und 73,33% auf Englisch geschrieben. Die Beziehung zwischen *Sprache* und *Externe FD* ist mit $\chi^2(9, n = 1252) = 39,46, p = 9,50 \cdot 10^{-6} < 0,01, \phi_C = 0,10$ auch statistisch hochsignifikant und schwach in Effektstärke.

Abhängigkeit Da *Sprache* und *Fakultät* auch miteinander korrelieren, wie in Abschnitt 4.2.2 gezeigt, muss hierbei beachtet werden inwiefern der Effekt von *Sprache* auf *Allgemeine FD* und *Externe FD* an *Sprache* liegt und wie viel der Effekt durch ihre Korrelation mit *Fakultät* bedingt ist. Siehe hierfür die Interaktion zwischen *Fakultät* und den verschiedenen FD-Publikationsarten in Abschnitt 4.2.6

4.2.6 Fakultäten und Forschungsdaten

Klassifikation Fakultäten Für PD-Dissertationen ist die relative und absolute Verteilung der Klassifikationsstufen für alle Dokumente sowie nach *Fakultät* × *Klassifikationsstufe* in Tabelle 4.9 gegeben. Dieselbe Aufteilung für alle Dissertationen ist in Tabelle B.2 gegeben.

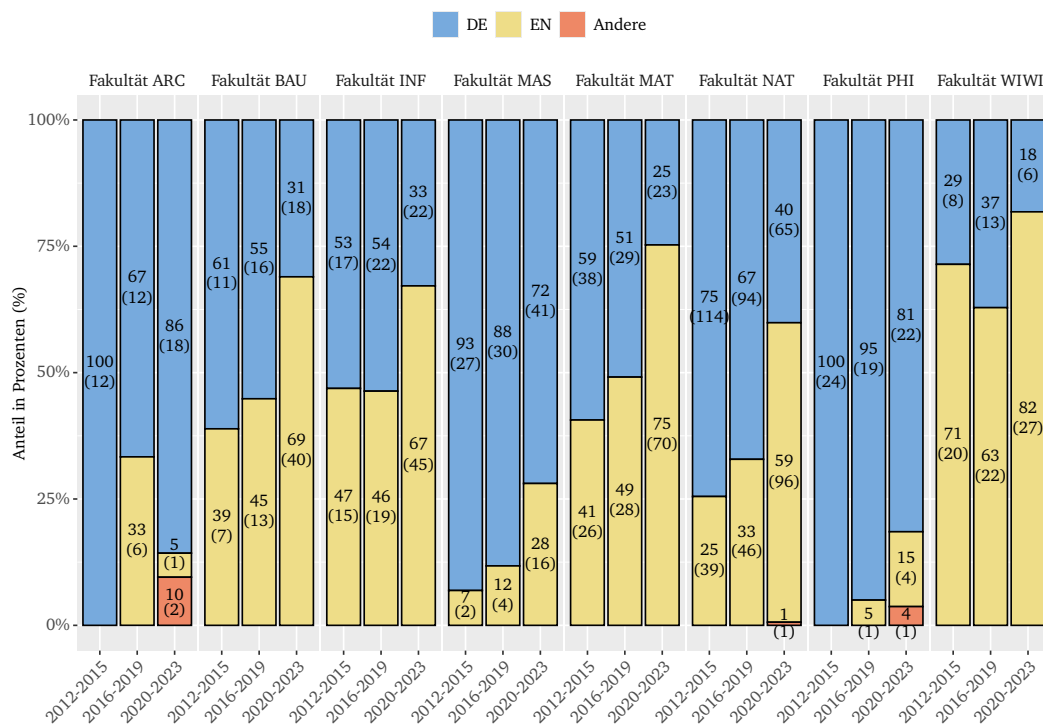


Abbildung 4.1: Sprachen der PD-Dissertationen nach Fakultät und Zeitgruppe. Die Höhe der Balken entsprechen dem relativen Anteil zur jeweiligen angepassten *Fakultät × Zeitgruppe × Sprache*-Gesamtanzahl. Absolute Werte in Klammern angeben.

Tabelle 4.9: FD-Klassifizierung der PD-Dissertationen aus der Stichprobe nach *Fakultät × Klassifikationsstufe* aufgegliedert. Angabe relativ zu der respektiven Gesamtanzahl für *Fakultät*. Absolute Werte in Klammern angeben.

	Stufe 1		Stufe 2		Stufe 3		Keine	
Fakultät ARC	15,69 %	(8)	0,00 %	(0)	54,90 %	(28)	29,41 %	(15)
Fakultät BAU	26,67 %	(28)	12,38 %	(13)	12,38 %	(13)	48,57 %	(51)
Fakultät INF	30,00 %	(42)	12,86 %	(18)	18,57 %	(26)	38,57 %	(54)
Fakultät JUR	—	(0)	—	(0)	—	(0)	—	(0)
Fakultät MAS	12,50 %	(15)	4,17 %	(5)	19,17 %	(23)	64,17 %	(77)
Fakultät MAT	15,42 %	(33)	4,67 %	(10)	22,90 %	(49)	57,01 %	(122)
Fakultät NAT	22,42 %	(102)	21,10 %	(96)	40,88 %	(186)	15,60 %	(71)
Fakultät PHI	22,54 %	(16)	0,00 %	(0)	33,80 %	(24)	43,66 %	(31)
Fakultät WIWI	8,33 %	(8)	1,04 %	(1)	25,00 %	(24)	65,63 %	(63)
Alle	20,14 %	(252)	11,43 %	(143)	29,82 %	(373)	38,61 %	(484)

Allgemeine FD

Die Interaktion für PD-Dissertationen zwischen *Fakultät* und *Allgemeine FD* ist mit $\chi^2(21, n = 1252) = 277,02, p = 1,47 \cdot 10^{-46} < 0,01, \phi_C = 0,27$ statistisch hochsignifikant mit einer fast moderaten Effektstärke. Für FD sind die respektiven Anteile pro Fakultät für *Stufe 1* ($\bar{x} = 19,19\%, s = 7,40\%P$), *Stufe 2* ($\bar{x} = 7,03\%, s = 7,64\%P$), *Stufe 3* ($\bar{x} = 28,45\%, s = 13,98\%P$) und *Keine* ($\bar{x} = 45,33\%, s = 17,32\%P$). Es variieren zwischen den Fakultäten also insbesondere *Stufe 3* und *Keine FD*, während *Stufe 1* und *Stufe 2* moderatere interfakultäre Variation erleben. Es besteht also ein vergleichsweise kleiner Kern an PD-Dissertationen, die in allen Fakultäten–zu moderat unterschiedlich großen Anteilen–FD aus *Stufe 1* oder *Stufe 2* publizieren, während der größte Unterschied zwischen Fakultäten dadurch bestimmt wird, wie groß die jeweilige Verteilung zwischen *Keine FD* und FD aus *Stufe 3* ist. Dieser Effekt kann in Abbildung 4.2, welche die relative und absolute Verteilung von FD für Fakultäten über die verschiedenen Zeitgruppen darstellt, gesehen werden. Mit $\chi^2(21, n = 1252) = 277,02,$

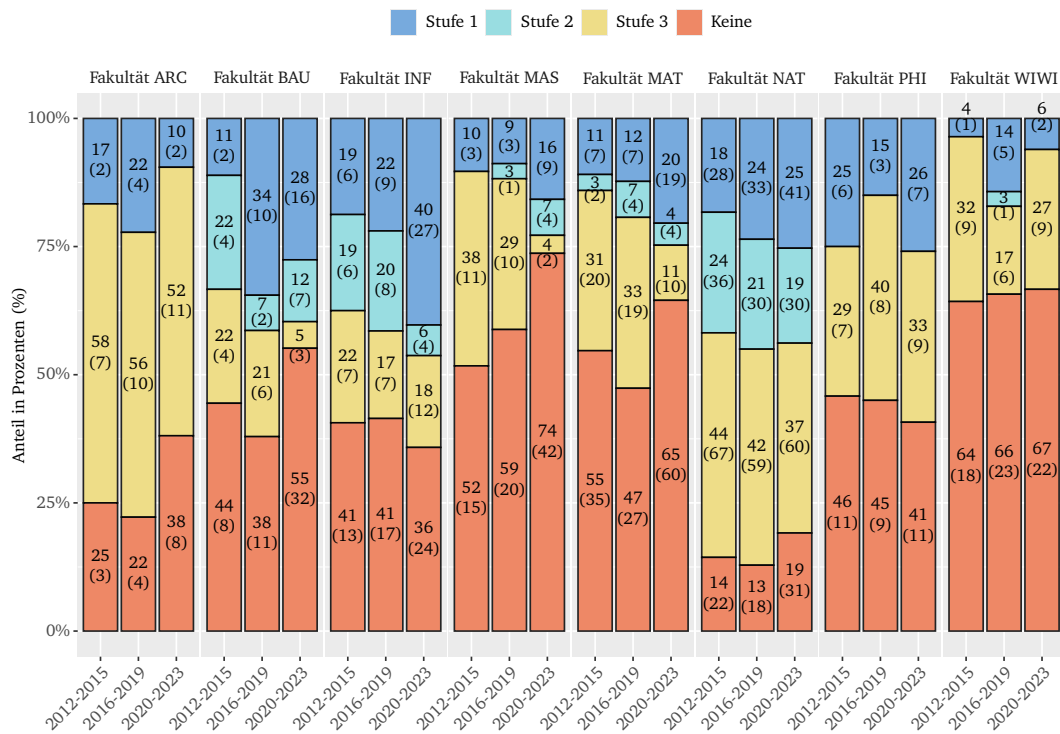


Abbildung 4.2: FD für PD-Dissertationen nach Fakultät, Zeitgruppe und Klassifikationsstufe. Die Höhe der Balken entsprechen dem relativen Anteil zur jeweiligen angepassten *Fakultät* × *Zeitgruppe* × *FD*-Gesamtanzahl. Absolute Werte in Klammern angegeben.

$p = 1,47 \cdot 10^{-46}, \phi_C = 0,27$ ist die Interaktion zwischen *Fakultät* und *Allgemeine FD* für PD-Dissertationen statistisch hochsignifikant mit einem Effekt fast moderater Stärke.

Signifikanz

Von den in Abbildung 4.2 dargestellten fakultätsspezifischen Interaktionen zwischen *Zeitgruppe* und *Allgemeine FD* waren Fakultät MAS ($\chi^2(6, n = 120) = 19,55, p =$

0,003, $\phi_C = 0,29$) und Fakultät MAT ($\chi^2(6, n = 214) = 16,30, p = 0,01, \phi_C = 0,20$) statistisch signifikant, wobei der Effekt für Fakultät MAS etwas stärker ist als für Fakultät MAT. Für die restlichen Fakultäten war die Interaktion statistisch nicht signifikant.

Integrierte Forschungsdaten

Die relative und absolute Verteilung von integrierten FD für Fakultäten über die verschiedenen Zeitgruppen hinweg ist in Abbildung 4.3 dargestellt. Mit $\chi^2(21, n =$

Integrierte FD

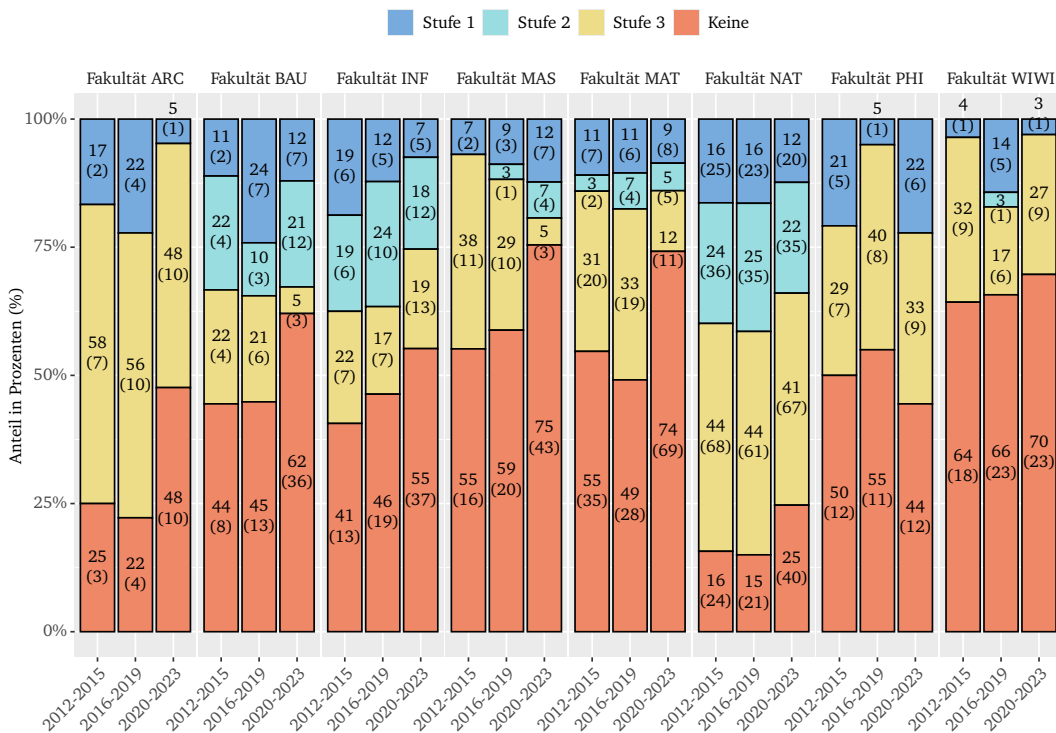


Abbildung 4.3: Integrierte FD für PD-Dissertationen nach Fakultät, Zeitgruppe und Klassifikationsstufe. Die Höhe der Balken entsprechen dem relativen Anteil zur jeweiligen angepassten Fakultät × Zeitgruppe × Integrierte FD-Gesamtanzahl. Absolute Werte in Klammern angegeben.

1252) = 263,00, $p = 9,94 \cdot 10^{-44}, \phi_C = 0,26$ ist die Interaktion zwischen Fakultät und Integrierten FD für PD-Dissertationen statistisch hochsignifikant mit einem Effekt fast moderater Stärke.

Von den in Abbildung 4.3 dargestellten fakultätsspezifischen Interaktionen zwischen Zeitgruppe und Allgemeine FD waren auch hier nur Fakultät MAS ($\chi^2(6, n = 120) = 16,98, p = 0,01, \phi_C = 0,27$) und Fakultät MAT ($\chi^2(6, n = 214) = 14,91, p = 0,02, \phi_C = 0,19$) statistisch signifikant, wobei der Effekt für Fakultät MAS etwas stärker ist als für Fakultät MAT. Hier ist feststellbar, dass die Interaktionen sowohl

Signifikanz

weniger signifikant wie auch schwächer sind, als die respektiven Interaktionen zwischen *Zeitgruppe* und *Integrierte FD*. Für die restlichen Fakultäten war die Interaktion statistisch nicht signifikant.

Begleitende Forschungsdaten

Begleitende FD Die relative und absolute Verteilung von begleitenden FD für Fakultäten über die verschiedenen Zeitgruppen hinweg ist in Abbildung 4.4 dargestellt. Mit $\chi^2(14, n =$

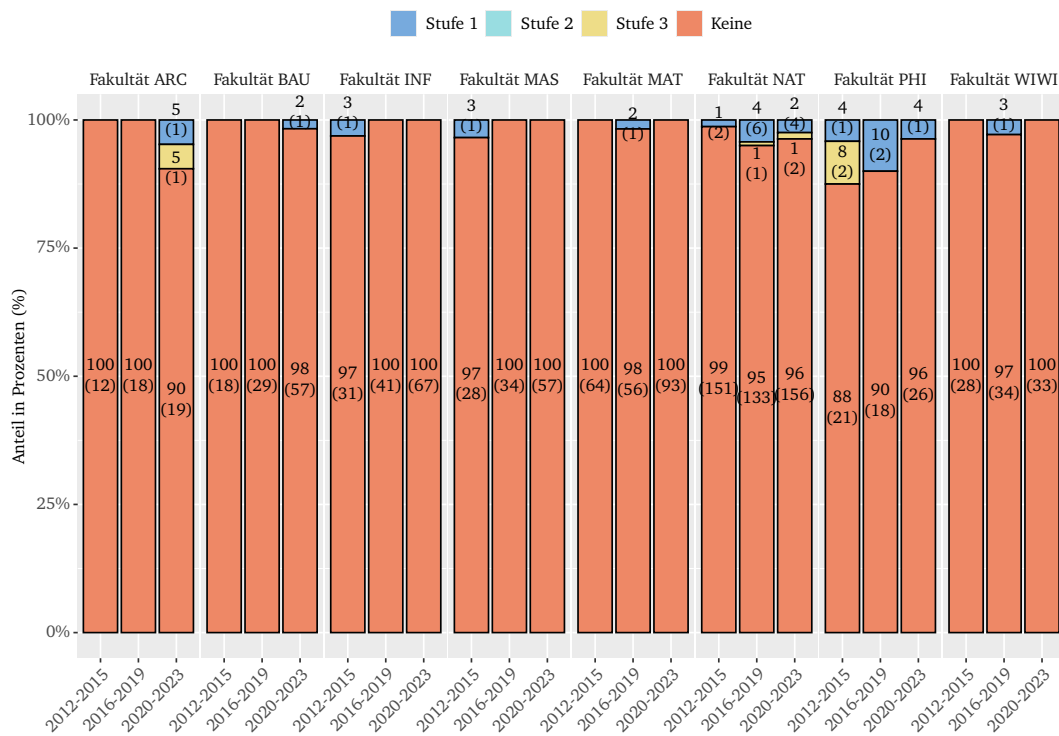


Abbildung 4.4: Externe FD für PD-Dissertationen nach Fakultät, Zeitgruppe und Klassifikationsstufe. Die Höhe der Balken entsprechen dem relativen Anteil zur jeweiligen angepassten *Fakultät × Zeitgruppe × Begleitende FD*-Gesamtanzahl. Absolute Werte in Klammern angegeben.

1 252) = 26,71, $p = 0,02$, $\phi_c = 0,10$ ist die Interaktion zwischen *Fakultät* und *Begleitende FD* für PD-Dissertationen statistisch signifikant mit einer schwachen Effektstärke.

Signifikanz Von den in Abbildung 4.4 dargestellten fakultätsspezifischen Interaktionen zwischen *Zeitgruppe* und *Begleitende FD* waren keine Interaktionen statistisch signifikant.

Externe Forschungsdaten

Die relative und absolute Verteilung von externen FD für Fakultäten über die verschiedenen Zeitgruppen hinweg ist in Abbildung 4.5 dargestellt. Mit $\chi^2(21, n = 1252) =$

Externe FD

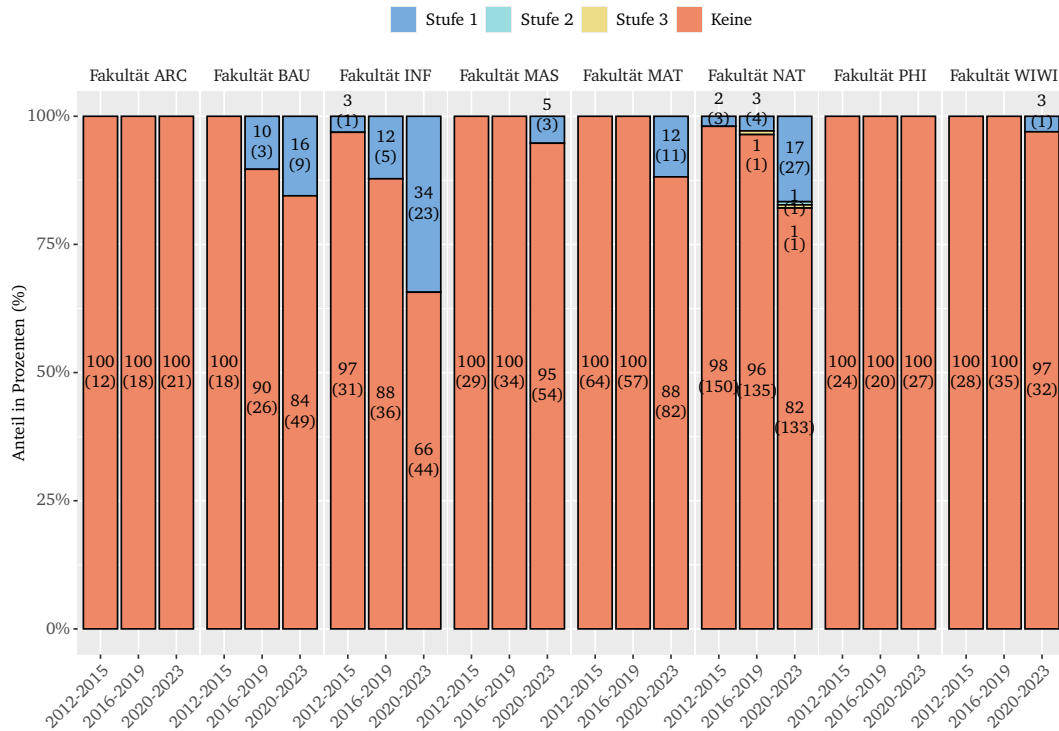


Abbildung 4.5: Externe FD für PD-Dissertationen nach Fakultät, Zeitgruppe und Klassifikationsstufe. Die Höhe der Balken entsprechen dem relativen Anteil zur jeweiligen angepassten Fakultät × Zeitgruppe × Externe FD-Gesamtanzahl. Absolute Werte in Klammern angegeben.

66,76, $p = 1,15 \cdot 10^{-6}$, $\phi_C = 0,13$ ist die Interaktion zwischen Fakultät und Begleitende FD für PD-Dissertationen statistisch hochsignifikant mit einer schwachen Effektstärke.

Von den in Abbildung 4.5 dargestellten fakultätsspezifischen Interaktionen zwischen Zeitgruppe und Allgemeine FD waren auch hier nur Fakultät INF ($\chi^2(2, n = 140) = 15,40, p = 0,0005, \phi_C = 0,33$) und Fakultät MAT ($\chi^2(2, n = 214) = 15,09, p = 0,001, \phi_C = 0,27$) statistisch signifikant, wobei der Effekt für Fakultät MAS etwas stärker ist als für Fakultät MAT.

Signifikanz

4.2.7 Externe Publikationen und Metadaten

Die Untersuchung von Metadatenangaben für externe Publikationen hat ergeben, dass keine der externen Publikationsformen in ihren Metadaten festhält, dass die Daten einer Dissertation aus dem LUH-Repo zugehörig sind. Dies ist zumindest im Teil auch

Externe Plattformen

dadurch bedingt, dass die Nutzung dedizierter FD-Repositorien bei den externen Publikationsmöglichkeiten in der Stichprobe in der Minderheit lag: Viele Daten wurden extern in dedizierten Datenbanken (z.B. Gensequenzen), auf GitHub / GitLab publiziert oder entsprachen bei kumulativen Dissertationen begleitenden Dateien, die in einem Journal zusammen mit der PDF-Datei hochgeladen wurden. Insgesamt wurden fast keine Plattformen genutzt, die von sich aus eine Kodifizierung entsprechender Metadaten ermöglichen.

LUH-Repo Zum Zeitpunkt dieser Arbeit wurde in den Metadaten des LUH-Repos nicht explizit verfasst, dass eine als verwandt markierte DOI den FD der Dissertation entspricht. Hier wurde dieselbe Relation verwendet, die auch genutzt wurde, um bei kumulativen Dissertationen auf die einzelnen Artikel zu verweisen, aus der die Dissertation besteht. Diese machten, mit erheblichem Abstand, den Großteil aller verknüpften DOIs aus; dies ist kongruent mit dem Ergebnis, dass die wenigsten externen FD auf dedizierte FD-Repositorien hochgeladen wurden–und dementsprechend auch keine DOI vergeben bekommen haben.

Datenmangel Durch die Kombination dieser Faktoren, konnten keine empirischen Daten zur Nutzung unterschiedlicher Metadatenstandards auf Basis des LUH-Repos und der sich darin befindenden Dissertationen gesammelt werden.

4.3 Diskussion

Abhängigkeit Die Ergebnisse in Abschnitt 4.2.2 zeigten, dass alle evaluierten Faktoren zu unterschiedlich starken–aber insgesamt eher schwachen–Graden voneinander abhängig sind. Dies verkompliziert die Analyse der jeweiligen Faktoren in Bezug auf deren Wirkung auf FD und deren unterschiedlichen Publikationsformen.

4.3.1 Sprache

Sprache Ein gutes Beispiel hierfür ist der Faktor *Sprache*. Die Annahme hinter diesem Faktor war, dass die Nutzung einer internationalen Sprache (i.e. *Englisch*) eventuell einen Einfluss auf die Beachtung und Einhaltung internationaler FDM-Standards haben könnte. Auf den ersten Blick scheint sich dies auch zu bewahrheiten, wie an zwei Beispielen gezeigt werden kann: So wurden ca. zwei Drittel aller PD-Dissertationen mit externen FD auf Englisch verfasst, obwohl die Mehrheit aller PD-Dissertationen auf Deutsch verfasst wurden, wie in Abschnitt 4.2.5 aufgezeigt wurde. Auch sinkt die Nutzungsrate von *Deutsch* und steigt für *Englisch* an je höher die Klassifikationsstufe einer PD-Dissertation ist–bis sie schließlich auf *Level 1* Parität erreichen. Beachtet man jedoch die Verteilung von *Sprache* in Abbildung 4.1 über die einzelnen Fakultäten, so zeigt sich, dass die Nutzung von *Englisch* für die meisten Fakultäten über die Jahre zugenommen hat. In demselben Zeitrahmen stieg auch die Nutzung von externen FD an, wie in der Verteilung von *Externe FD* über die verschiedenen Fakultäten und Zeitgruppen hinweg in Abbildung 4.5 gezeigt wurde.

Fakultärer Sprachanteil Dazu kommt, dass in einigen Fakultäten, die vergleichsweise viel PD-Dissertationen

der dritten Stufe generieren, wie z.B. in Fakultät ARC, der Anteil an englischsprachigen PD-Dissertationen sehr gering ist und auch nicht über die Jahre zugenommen hat, was die relative Verteilung von *Stufe 3* unter den zwei Sprachen maßgeblich beeinflusst.

Insgesamt ergibt sich für *Sprache* das Bild, dass der Faktor nur indirekt mit denen der FD korreliert. Es erscheint maßgeblicher, wie *Sprache* durch *Fakultät* geprägt wird und wie die Fakultäten unterschiedliche Publikationsverhalten besitzen (wie in Abschnitt 4.2.6 gezeigt wurde).

Schluss

4.3.2 Externe Forschungsdaten

Besonders interessant ist das Publikationsverhalten zu externen FD–auch da diese die eigentlich ideale Publikationsform darstellen. Betrachtet man Abbildung 4.5, so ist deutlich erkennbar, dass nur drei der neun Fakultäten für den Hauptteil aller extern publizierten FD verantwortlich sind und in mehr als einer Zeitgruppe externe FD produzierten: Fakultät BAU, Fakultät INF und Fakultät NAT. Von diesen nutzt insbesondere Fakultät INF externe FD zu einem besonders hohen Anteil. Jedoch nutzen, auch von diesen Fakultäten, nur die wenigsten dedizierte FD-Repositoryen. Stattdessen sind die meisten externen Publikationen Programmiercode und Skripte auf GitHub / GitLab (für Fakultät INF), in Datenbanken eingereichte Gensequenzen (für Fakultät NAT) oder Beirichungen zu Journal-Artikeln für kumulative Dissertationen. Das bedeutet, dass der Anteil an extern publizierten FD zwar insgesamt zunimmt, die Nutzung von entsprechenden Fachrepositorien bei Dissertationen jedoch weiterhin verschwindend gering bleibt. Tatsächlich war der Anteil so gering, dass es keine externe FD Veröffentlichung gab, in deren Metadaten spezifiziert wurde, dass diese dem Werk einer Dissertation zugehörig sei, was es unmöglich machte, empirisch zu untersuchen, welche Metadateneinträge hierfür am häufigsten genutzt werden. Insgesamt besteht dennoch ein Trend dazu, besonders hochwertige Daten eher extern zu publizieren, statt diese in der Dissertation integriert zu veröffentlichen. Dies lässt sich daran erkennen, dass der Gesamtanteil an allgemeinen und externen *Stufe 1* PD-Dissertationen über die Zeitgruppen stetig zugenommen hat, gleichzeitig aber für integrierte FD abgenommen hat, wie in Tabelle 4.7 gezeigt wurde.

Externe FD

4.3.3 Zeitliche Entwicklung von Forschungsdaten und Fakultäten

Abgesehen davon, dass extern publizierte FD über die Jahre zugenommen haben, lassen sich einige andere allgemeine Trends für PD-Dissertationen im LUH-Repo erkennen. Ein auf den ersten Blick kontraintuitiver Trend ist, dass der Anteil an allgemeinen PD-Dissertationen ohne publizierte FD für die letzte Jahresgruppe 2020–2023* stark zugenommen hat. Betrachtet man jedoch die allgemeine Nutzungsrate des LUH-Repos nach Fakultäten und Zeitgruppen (Tabelle 4.4), lässt sich feststellen, dass sich die Nutzungsrate für die letzte Zeitgruppe besonders stark verändert hat: Insbesondere in der vorherigen Zeitgruppe schwach vertretene Fakultäten wie Fakultät BAU, Fakultät INF und Fakultät PHI haben ihren Anteil an LUH-Repo-Dissertationen stark erhöht. Auch Fakultät MAS, welche besonders konsistent wenige FD veröffentlicht, hat seinen Anteil

Zeitliche Trends

um ca. 65% gegenüber der vorherigen Zeitgruppe erhöht. Während diese Kompositionsveränderung des LUH-Repos den abnehmenden FD-Trend zumindest zum Teil erklären, zeigt Abbildung 4.2 dennoch auf, dass dieser Trend auch einige Fakultäten intern betrifft: Nämlich Fakultät ARC, Fakultät BAU, Fakultät JUR und Fakultät MAS (und in deutlich geringerem Umfang auch Fakultät NAT und Fakultät WIWI). Was genau diese Abnahme verursacht ist nicht bekannt und Bedarf weiterer Beobachtung und Erforschung.

Begleitende FD

Was die Option betrifft, FD begleitend zu der Dissertation im LUH-Repo zu veröffentlichen, so lässt sich hierzu sagen, dass es eine sehr wenig genutzte Möglichkeit ist. Wie in Abbildung 4.4 gezeigt wurde, gibt es allerdings einen statistisch signifikanten kleinen aber konsistenten Kern an Nutzern, die besonders häufig der Fakultät NAT und der Fakultät PHI zugeordnet werden können. Es erscheint unwahrscheinlich, dass die Nutzungsrate dieser Publikationsart sich über die nächsten Jahre ändern wird, da zumindest bisher *Zeit* kein statistisch signifikanter Faktor war (siehe Abschnitt 4.2.4).

4.3.4 Handlungsempfehlungen

*Empfehlungs-
kategorien*

Die Resultate in Abschnitt 4.2 zeigen mehrere Möglichkeiten auf, was die TIB / die LUH tun oder beachten kann, um ihre Services weiter auszubauen. Diese Möglichkeiten teilen sich in zwei Hauptbereiche: Nutzung des LUH-Repos und verbessertes FDM durch die Promovierenden.

Nutzung des LUH-Repos

Nutzergruppen

Wie in Abschnitt 4.2.1 dargestellt, wird das LUH-Repo von den verschiedenen Fakultäten der LUH unterschiedlich stark für die Publikation von Dissertationen genutzt. Während Promovierende aus sieben der neun Fakultäten inzwischen mindestens annähernd die Hälfte aller ihrer neuen Dissertationen auf LUH-Repo publizieren, nämlich die Intermediär- und Intensivnutzer des Repositoriums, gibt es hierbei zwei bedeutende Ausreißer: Die Geringnutzer des LUH-Repo, Fakultät JUR und Fakultät MAS, welche weniger als ein Viertel ihrer neuen Dissertationen im LUH-Repo publizieren. Dieser Befund stimmt mit früherer Forschung überein, da diese zeigte, dass sowohl die Ingenieurwissenschaften (und verwandte Unterdisziplinen) wie auch die Rechtswissenschaften besonders schwach im Open Access Bereich vertreten sind [68, 69, 70, 71, 72]. Dennoch bedeuten diese geringen Werte, dass hier aktiver Handlungsbedarf seitens der LUH und der TIB besteht. So sollte vermehrt Aufklärungs-, Beratungs- und Werbearbeit auf Fakultät JUR und Fakultät MAS ausgerichtet werden.

*Rechtswissen-
schaft*

Hier kann insbesondere im Bereich der Rechtswissenschaft erfolgreiche Aufklärungsarbeit stattfinden, da Rechtswissenschaftler sich häufig nicht der Möglichkeit von Open Access bewusst sind—so sind z.B. die wenigsten mit ihrem Zweitveröffentlichungsrecht vertraut [73, S. 91]. Gleichzeitig ist in der Rechtswissenschaft die Verbreitung von Dissertationen in Deutschland „äußerst mangelhaft“ [74, S. 50], wobei ein vermehrter Wandel auf Open Access Abhilfe schaffen könnte.

Kontakt

Allgemein sollte zu diesen beiden Fakultäten Kontakt aufgebaut werden und jeweils

gemeinsam besprochen werden, welche Bedenken vor einer höheren Nutzung des LUH-Repos bestehen. Hierbei sollten die diversen Vorteile von Open Access erörtert [27] und auch aufgezeigt werden, dass z.B. ein bisheriger Mangel an High Impact Open Access Publikationen kein Indikator dafür ist, dass ein entsprechender Markt nicht sehr schnell entwickelt werden kann [75].

Forschungsdaten

Abschnitt 4.2 hat gezeigt, dass nur ein Drittel aller Autoren von PD-Dissertationen aus dem LUH-Repo FD der *Stufe 1* und *Stufe 2* publiziert hat, während ein weiteres Drittel überhaupt keine FD publiziert hat. Unter Ausschluss integrierter FD stieg hierbei der Anteil jener, die keine FD publizieren, sogar auf über 90 %.

*Geringe
FD-Zahlen*

Entsprechend gibt es hier zwei separate aber verwandte Bereiche, bei denen Promovierende an der LUH eventuell zusätzliche Unterstützung erfahren oder durch vermehrte Werbearbeit auf existierende Angebote der LUH / TIB aufmerksam gemacht werden sollten: (i) die Identifikation von eigenen PDs, die zur Verfügung gestellt werden sollten, und (ii) die Möglichkeit, wie dies durch externe Repositorien nach den FAIR-Prinzipien stattfinden kann.

*Empfehlungs-
bereiche*

Ersteres kann insofern ein wichtiger Faktor sein, da die FD-Richtlinie der LUH eigentlich vorsieht, dass die FD nach Möglichkeit publiziert werden sollten [76]. Dies scheint aber nicht in vollem Umfang durch die Promovierenden umgesetzt zu werden. Einen Grund für diesen Umstand könnte sein, dass Promovierende einfach unterschätzen, welche PDs für die Veröffentlichung relevant sein könnten.

*Identifikation
eigener PDs*

Eine Möglichkeit, wie für das bestehende Beratungs- und Informationsangebot zu FDM unter den Studierenden und Promovierenden Werbearbeit gemacht und gleichzeitig auch erweitert werden kann, wäre die Nutzung eines künstlichen Intelligenz (KI) Modells, welches darauf trainiert wurde, zu erkennen, welche Datensätze in einer Arbeit implizit genutzt werden und daher eventuell auch explizit veröffentlicht werden sollten. Eine Möglichkeit hierzu wäre das *DataSeer*-Projekt [77], welches entwickelt wurde, um Forschenden und Herausgebern von wissenschaftlichen Journalen dabei zu helfen, zu identifizieren, welche Datensätze einem Artikel zugrunde liegen und welche daher idealerweise auch geteilt oder referenziert werden sollten. Da es sich bei *DataSeer* um ein quelloffenes Projekt handelt, wäre hier eine Möglichkeit, dass die TIB oder die LUH eine eigene Instanz hostet, welche LUH-Angehörigen zur Verfügung steht. Hier könnte dann durch Plakate und andere Medien eine „Check Your Data“-Kampagne gestartet werden, welche den KI-Aspekt des Projektes in den Vordergrund stellt, um die Neugier der Studierenden und Promovierenden dafür zu nutzen, deren wissenschaftliche Praxis oder zumindest wissenschaftliche FD-Achtsamkeit zu verbessern.

*DataSeer-
Kampagne*

Was die Möglichkeit betrifft, FD in ein externes Repository hochzuladen, so könnte ein mangelndes Wissen, welche Repositorien existieren und wie diese genutzt werden können, ein Faktor sein, warum so wenige PD-Dissertationen im LUH-Repo externe FD in dedizierte FD-Repositorien hochgeladen haben. Hier sollte ver-

Repositorien

mehrt Werbearbeit für das institutionelle FD-Repositorium und für die allgemeine FDM-Informationswebseite der LUH gemacht werden–welche bereits auf geeignete Repositorien-Suchmöglichkeiten wie z.B. *re3data* hinweist [78].

Prominenz Eine Option hierfür wäre, dass die Ressourcen zu externen Beratungsmöglichkeiten prominenter dargestellt werden. So könnte ein Link zu der entsprechenden Seite auf der Hauptseite des Informationsportals eingestellt werden.

Pflichtkurse Eine andere Option hierfür wäre, den bereits existierenden autodidaktischen FDM-Einführungskurs der LUH für Studierende und Promovierende verpflichtend zu machen, sodass diese einen Nachweis hierzu erbringen müssen. Für Studierende müsste dies in den Semestern vor der Abschlussarbeit und für Promovierende in den ersten Semestern ihrer Forschungstätigkeit der Fall sein, um die gewünschte Wirkung zu maximieren. Auch könnte, unter Absprache mit den Fakultäten und der Universitätsleitung, eine verpflichtende Erstellung eines DMP einerseits die wissenschaftliche Praxis im Allgemeinen verbessern und andererseits den Studierenden bzw. Promovierenden die hohe Relevanz der Thematik FDM vermitteln.

Kapitel 5

Schlussfolgerungen

Die einzelnen Forschungsfragen dieser Abschlussarbeit lauteten, wie in Kapitel 1 aufgelistet, wie folgt:

*Forschungs-
fragen*

1. Inwiefern wird der Umgang mit FD für Promotionsvorhaben in Deutschland bereits in verbindlichen verwaltungsrechtliche Dokumente geregelt?
 - a) Inwiefern wird der Umgang mit FD in den allgemeinen Richtlinien einer Institution geregelt, welche auch Promovierende und ihr Forschungsvorhaben betreffen?
 - b) Inwiefern wird der Umgang mit FD in Promotions- und Prüfungsordnungen geregelt?
2. Auf welche Art und Weise haben Promovierende, welche ihre Dissertationen 2012–2023 in einem institutionellen Repositorium publiziert haben, ihre FD in ihre Publikation integriert?
 - a) Für welchen Anteil an Dissertationen wurden FD in welcher Form nach dem Klassifikationssystem von REILLY u. a. [3] veröffentlicht?
 - b) Inwiefern hat sich das Publikationsverhalten zu FD in Dissertationen in den letzten zwölf Jahren verändert?
 - c) Inwiefern unterscheiden sich die verschiedenen wissenschaftlichen Fachrichtungen in ihrem Publikationsverhalten in Bezug auf FD aus Dissertationen?
 - d) Wie wird in den Metadaten von FD aus Dissertationen sichtbar gemacht, dass es eine dazugehörige Dissertation gibt?
 - e) Wie wird in den Metadaten von Dissertationen sichtbar gemacht, dass es dazugehörige FD gibt?

Der Kontext der zweiten Forschungsfrage beschränkte sich dabei, aus Gründen der Relevanz und des Umfangs dieser Abschlussarbeit, jeweils auf Dissertationen und auf das LUH-Repo.

Die erste Forschungsfrage, wie auch ihre spezifischeren Unterforschungsfragen, wurde in Kapitel 3 behandelt.

*Erste For-
schungsfrage
Allgemeine
Richtlinien*

Für die erste Unterfrage, die sich auf allgemeine Richtlinien bezieht, wurde durch eine quantitative Auswertung einer repräsentativen Stichprobe ($n = 115$) aller promotionsberechtigter Institutionen in Deutschland ($n = 163$) gezeigt, dass 44,35 % ($n = 51$) der Institutionen eine dedizierte FD-Richtlinie besitzen. Weitere 43,48 % ($n = 50$) besitzen hierbei zumindest eine Verpflichtung zum Einhalten der Regeln der GWP.

Hierbei wurden jeweils statistisch signifikante Abhängigkeiten zwischen der Trägerschaft und dem Typ der Forschungsinstitution zu der Richtlinien-Klassifikationsstufe der Institution festgestellt, jedoch nicht mit dem Bundesland, in dem sich die Institution befindet. Diese Ergebnisse aus Kapitel 3 sind konform mit den bisher bekannten Daten zu FD-Richtlinien in Deutschland, wie sie in Kapitel 2 zusammengefasst wurden. Zusätzlich zeigen die Ergebnisse auf, dass, zumindest bisher, keine Homogenisierung der FD-Richtlinien-Bezeichnungen in Deutschland stattgefunden hat.

*Spezifische
Richtlinien*

Für die zweite Unterfrage, welche sich auf promotionspezifische Dokumente bezieht, wurde gezeigt, dass nur 13,91 % ($n = 16$) aller Stichproben-Institutionen mindestens eine Promotions- oder Prüfungsordnung besitzen, die verpflichtende Richtlinien zum Umgang mit FD enthält. Weitere 53,04 % ($n = 61$) enthalten jedoch verpflichtende Richtlinien zum Einhalten der GWP. Es wurden statistisch signifikante Abhängigkeiten zwischen der Richtlinien-Klassifikationsstufe für promotionspezifische Dokumente der Institution und dem Bundesland, dem Institutionstyp und der höchsten Klassifikationsstufe für allgemeine Richtlinien der Institution gefunden.

Antwort

Hiermit konnte die erste Forschungsfrage quantitativ beantwortet werden. In Deutschland haben fast die Hälfte aller promotionsberechtigter Institutionen eine allgemeine (und verpflichtende) FD-Richtlinie, während nur die Wenigsten auch FD oder FDM in den Promotions- bzw. Prüfungsordnungen erwähnen. Für Details zu den einzelnen Ergebnissen und Verhältnissen wird auf Abschnitte 3.2 und 3.3 verwiesen.

*Zweite For-
schungsfrage*

Die zweite Forschungsfrage, wie auch ihre spezifischeren Unterforschungsfragen, wurde in Kapitel 4 behandelt.

*Publikationsar-
ten*

Für die erste Unterfrage, wurde gezeigt, dass ca. 61,34 % ($n = 768$) aller PD-Dissertationen ($n = 1252$) aus der mehrschichtigen Stichprobe ($n = 1441$) zumindest einen Teil ihrer FD veröffentlicht haben. Von den PD-Dissertationen haben hierbei 57,03 % ($n = 714$) zumindest einen Teil ihrer FD in der Dissertation integriert, 1,76 % ($n = 22$) als begleitende Dateien und 7,43 % ($n = 93$) auf einer externen Plattform veröffentlicht.

*Zeitliche
Entwicklung*

Bei der zweiten Unterfrage, welche die zeitliche Entwicklung betrifft, konnte festgestellt werden, dass sich das FD-Publikationsverhalten über einen Zeitraum von zwölf Jahren statistisch signifikant geändert hat und vor allem der Anteil an externen Publikationen in den letzten vier Jahren stark zugenommen hat.

Fakultäten

Für die dritte Unterfrage, welche die Fakultäten betrifft, konnte festgestellt werden, dass es statistisch signifikante Unterschiede zwischen den Fakultäten in ihrem FD-Publikationsverhalten gibt. So zeichnen sich insbesondere Fakultät INF, Fakultät BAU und Fakultät NAT dadurch aus, einen höheren Anteil an externen Publikationen zu besitzen, als die anderen Fakultäten. Ein weiterer interessanter Unterschied zwischen den Fakultäten ist auch, dass diese zu unterschiedlich großen Anteilen das LUH-Repo nutzen, um Dissertationen zu veröffentlichen. So lassen sich die Fakultäten hierfür in drei Gruppen einteilen: Geringnutzer (Fakultät JUR und Fakultät MAS), Intermediärnutzer (Fakultät BAU, Fakultät INF und Fakultät PHI) und Intensivnutzer (Fakultät ARC, Fakultät MAT, Fakultät NAT und Fakultät WIWI).

Für die vierte und fünfte Unterfrage, welche die Metadaten betreffen, wurde festgestellt, dass keine der PD-Dissertationen auf der externen FD-Plattform in den dortigen Metadaten festhält, dass die FD mit einer Dissertation verknüpft sein sollten. Dies ist aber, zumindest teilweise, dadurch bedingt, dass die wenigsten externen Publikationen auf dedizierten FD-Repositoryn veröffentlicht wurden, sondern auf Plattformen, die keine Funktion für derlei Metadaten besitzen (z.B. *GitHub* oder *GitLab*). Von Seiten der Dissertationen aus, wiederum, wurde nur eine allgemeine Relation genutzt, aus der nicht ersichtlich ist, dass es sich dabei um FD handelt—stattdessen wird diese Relation hauptsächlich dafür genutzt, Artikel einer kumulativen Dissertation zu referenzieren.

Metadaten

Somit konnte auch die zweite Forschungsfrage weitestgehend quantitativ beantwortet werden. Die Dissertationen aus dem LUH-Repo produzieren und veröffentlichen zum Großteil mindestens einen Anteil ihrer FD. Dabei sind die meisten FD aber auf eine Art und Weise veröffentlicht, die sie nur schwer für weitere Forschung wiederverwendbar macht. Insgesamt, lässt sich zusammenfassen, dass nur die wenigsten FD denFAIR-Prinzipien entsprechen. Dies deutet an, dass Promovierende in diesem Bereich vermehrt Informationen und Beratung benötigen. Entsprechende Vorschläge, wie dies bewerkstelligt werden könnte, wurden in Abschnitt 4.3 gegeben. Für Details zu den einzelnen Ergebnissen und Verhältnissen wird auf Abschnitte 4.2 und 4.3 verwiesen.

Antwort

DIESE SEITE WURDE ABSICHTLICH LEER GELASSEN

Literaturverzeichnis

- [1] WILKINSON, Mark D.; Michel DUMONTIER; IJsbrand Jan AALBERSBERG; Gabrielle APPLETON; Myles AXTON u. a. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, Bd. 3 (2016), Nr. 1, Artikelnummer 160018. ISSN 2052-4463. Verfügbar unter: <https://doi.org/10.1038/sdata.2016.18>.
- [2] TENOPIR, Carol; Sanna TALJA; Wolfram HORSTMANN; Elina LATE; Dane HUGHES u. a. Research Data Services in European Academic Research Libraries. *LIBER Quarterly: The Journal of the Association of European Research Libraries*, Bd. 27 (2017), Nr. 1, S. 23–44. Verfügbar unter: <https://doi.org/10.18352/lq.10180>.
- [3] REILLY, Susan; Wouter SCHALLIER; Sabine SCHRIMPF; Eefke SMIT und Max WILKINSON. *Opportunities of Data Exchange: Report on Integration of Data and Publications*. 2011. Verfügbar unter: <https://hdl.handle.net/10013/epic.40198.d001>.
- [4] PIWOWAR, Heather A. und Todd J. VISION. Data reuse and the open data citation advantage. *PeerJ*, Bd. 1 (2013), e175. ISSN 2167-8359. Verfügbar unter: <https://doi.org/10.7717/peerj.175>.
- [5] WEISBROD, Dirk; Ben KADEN und Michael KLEINEBERG. eDissPlus – Optionen für die Langzeitarchivierung dissertationsbezogener Forschungsdaten aus Sicht von Bibliotheken und Forschenden. In: *{E-Science-Tage: Forschungsdaten managen}*. Humboldt-Universität zu Berlin, 2017., S. 189–198. Verfügbar unter: <https://doi.org/10.18452/22310>.
- [6] KADEN, Ben und Michael KLEINEBERG. Zur Veröffentlichung dissertationsbezogener Forschungsdaten: Perspektiven und Kompetenzen von Promovierenden an Berliner Universitäten. *Bausteine Forschungsdatenmanagement*, 2018-10, Nr. 1, S. 64–69. Verfügbar unter: <https://doi.org/10.17192/bfdm.2018.1.7938>. Verfügbar unter: <https://bausteine-fdm.de/article/view/7938>.
- [7] WEISBROD, Dirk. Pflichtablieferung von Dissertationen mit Forschungsdaten an die DNB – Anlagerungsformen und Datenmodell. *o-bib. Das offene Bibliotheksjournal*, Bd. 5 (2018), Nr. 2, S. 72–78. Verfügbar unter: <https://doi.org/10.5282/o-bib/2018H2S72-78>.
- [8] DEUTSCHE NATIONALBIBLIOTHEK [DNB]. *Policy der Deutschen Nationalbibliothek für dissertationsbezogene Forschungsdaten*. Deutsche Nationalbibliothek, 2017. Verfügbar unter: <https://d-nb.info/114060242X/34>.

- [9] WÜNSCHE, Stephan. *Forschungsdaten in musikwissenschaftlichen und musikpädagogischen Dissertationen: Eine Stichprobe anhand der im Jahr 2015 in Deutschland angenommenen Arbeiten*. Masterarbeit. Humboldt-Universität zu Berlin, Philosophische Fakultät, 2018. Verfügbar unter: <https://doi.org/10.18452/19457>.
- [10] SCHÖPFEL, Joachim; Južnič PRIMOŽ; Hélène PROST; Cécile MALLERET; Ana ČEŠAREK u. a. *Dissertations and Data*. In: {*GL17 International Conference on Grey Literature*. Amsterdam, Niederlande: 2015-12.}, S. 1–33. Verfügbar unter: <https://hal.science/hal-01285304v1>.
- [11] MARTIN, Christiane Laura. *Wissenschaftliche Bibliotheken als Akteure im Forschungsdatenmanagement*. *LIBREAS. Library Ideas: Forschungsdaten. Metadaten. Noch mehr Daten. Forschungsdatenmanagement*, Bd. 23 (2013), S. 12–21. ISSN 1860-7950. Verfügbar unter: <https://doi.org/10.18452/9039>.
- [12] PAMPEL, Heinz; Paul VIERKANT; Frank SCHOLZE; Roland BERTELMANN; Maxi KINDLING u. a. *Making Research Data Repositories Visible: The re3data.org Registry*. *PLOS ONE*, Bd. 8 (2013), Nr. 11, e78080. Public Library of Science. Verfügbar unter: <https://doi.org/10.1371/journal.pone.0078080>.
- [13] KINDLING, Peter, Maxi und Schirmbacher. „Die digitale Forschungswelt“ als Gegenstand der Forschung: Lehrstuhl Informationsmanagement. *Information – Wissenschaft – Praxis*, Bd. 64 (2013), 2–3, S. 127–136. Verfügbar unter: <https://doi.org/10.1515/iwp-2013-0017>.
- [14] DEUTSCHE FORSCHUNGSGEMEINSCHAFT [DFG]. *Leitlinien zum Umgang mit Forschungsdaten*. 2015. Verfügbar unter: <https://www.dfg.de/resource/blob/172112/4ea861510ea369157afb499e96fb359a/leitlinien-forschungsdaten-data.pdf>.
- [15] HOCHSCHULREKTORENKONFERENZ [HRK]. *Management von Forschungsdaten – eine zentrale strategische Herausforderung für Hochschulleitungen: Empfehlung der 16. Mitgliederversammlung der HRK am 13. Mai 2014 in Frankfurt am Main*. 2014. Verfügbar unter: https://www.hrk.de/fileadmin/_migrated/content_uploads/HRK_Empfehlung_Forschungsdaten_13052014_01.pdf.
- [16] RAT FÜR INFORMATIONENINFRASTRUKTUREN [RFII]. *Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland*. Göttingen: 2016-05. Verfügbar unter: <http://nbn-resolving.de/urn:nbn:de:101:1-201606229098>.
- [17] DEUTSCHE FORSCHUNGSGEMEINSCHAFT [DFG]. *Förderung von Informationsinfrastrukturen für die Wissenschaft: Ein Positionspapier der Deutschen Forschungsgemeinschaft*. 2018. Verfügbar unter: http://www.dfg.de/download/pdf/foerderung/programme/lis/positionspapier_informationsinfrastrukturen.pdf.

- [18] GEMEINSAME WISSENSCHAFTSKONFERENZ [GWK]. *Bund-Länder-Vereinbarung zu Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur (NFDI)*. 2018-11. Verfügbar unter: <https://www.gwk-bonn.de/fileadmin/Redaktion/Dokumente/Papers/NFDI.pdf>.
- [19] NATIONALE FORSCHUNGSDATENINFRASTRUKTUR. *Satzung*. 2020-11. Verfügbar unter: <https://www.nfdi.de/wp-content/uploads/2021/05/Satzung-NFDI-eV.pdf>.
- [20] BUNDESMINISTERIUM FÜR BILDUNG UND FORSCHUNG [BMBF]. *Eckpunkte BMBF Forschungsdatengesetz*. 2024-02. Verfügbar unter: https://www.bmbf.de/SharedDocs/Downloads/de/2024/240306_eckpunktepapier-forschungsdaten.pdf.
- [21] DEUTSCHE FORSCHUNGSGEMEINSCHAFT [DFG]. *Die Wissenschaft braucht ein Forschungsdatengesetz!: Positionierung der Deutschen Forschungsgemeinschaft (DFG)*. Version v1. Zenodo, 2023-04. Verfügbar unter: <https://doi.org/10.5281/zenodo.8223970>.
- [22] RAT FÜR INFORMATIONENINFRASTRUKTUREN [RFII]. *Anforderungen an die Ausgestaltung eines Forschungsdatengesetzes und eines Dateninstituts*. Göttingen: 2023-07. Verfügbar unter: <https://rfii.de/download/10-07-2023-pressemitteilung-diskussionsimpuls-zum-forschungsdatengesetz-und-dateninstitut/>.
- [23] KADEN, Ben. Wissenschaftliche Bibliotheken als Akteure im Forschungsdatenmanagement. *LIBREAS. Library Ideas*, Bd. 33 (2018), S. 1–8. ISSN 1860-7950. Verfügbar unter: <https://doi.org/10.18452/19284>.
- [24] CAMPBELL, Hamish A.; Mariana A. MICHELI-CAMPBELL und Vinay UDYAWER. Early Career Researchers Embrace Data Sharing. *Trends in Ecology & Evolution*, Bd. 34 (2019), Nr. 2, S. 95–98. Verfügbar unter: <https://doi.org/10.1016/j.tree.2018.11.010>.
- [25] DEUTSCHE FORSCHUNGSGEMEINSCHAFT [DFG]. *Vorschläge zur Sicherung guter wissenschaftlicher Praxis: Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“*. Weinheim: Wiley-VCH, 2013. ISBN 978-3-527-67918-8. Verfügbar unter: <https://doi.org/10.1002/9783527679188.oth1>.
- [26] HOPF, David; Sarah DELLMANN; Christian HAUSCHKE und Marco TULLNEY. *Wirkungen von Open Access. Literaturstudie über empirische Arbeiten 2010-2021*. Hannover: Technische Informationsbibliothek [TIB], 2022. Verfügbar unter: <https://doi.org/10.34657/7666>.
- [27] BAUTISTA-PUIG, Nuria; Carmen LOPEZ-ILLESCAS; Felix de MOYA-ANEGON; Vicente GUERRERO-BOTE und Henk F. MOED. Do journals flipping to gold open access show an OA citation or publication advantage? *Scientometrics*, Bd. 124 (2020), Nr. 3, S. 2551–2575. ISSN 1588-2861. Verfügbar unter: <https://doi.org/10.1007/s11192-020-03546-x>.

- [28] GRAY, Jim. *Data Quality Pyramid*. 2009-04. Verfügbar unter: <https://www.slideshare.net/JamesWGray/data-quality-pyramid>.
- [29] KHAN, Aasif Mohammad; Fayaz Ahmad LOAN; Umer Yousuf PARRAY und Sozia RASHID. *Information Discovery and Delivery*, Bd. 52 (2024), Nr. 1. Emerald Insight. ISSN 2398-6247. Verfügbar unter: <https://doi.org/10.1108/IDD-07-2022-0069>.
- [30] RE3DATA.ORG. 2024-06. Suchanfrage für alle institutionellen Repositorien aus Deutschland. Verfügbar unter: <https://www.re3data.org/search?query=&types%5B%5D=institutional&countries%5B%5D=DEU>. [Zugriff am 2024-06-06]. Suchergebnis archiviert unter <https://www.doi.org/10.5281/zenodo.11401021>.
- [31] DUBLIN CORE™ METADATA INITIATIVE. *DCMI Metadata Terms*. 2020-01. Verfügbar unter: <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>. [Zugriff am 2024-06-06].
- [32] DATACITE METADATA WORKING GROUP. *DataCite Metadata Schema for the Publication and Citation of Research Data and Other Research Outputs*. Version 4.5. DataCite e.V., 2024-03. Verfügbar unter: <https://doi.org/10.14454/g8e5-6293>.
- [33] RE3DATA.ORG. 2024-06. Suchanfrage für alle Repositorien mit Informationen zu einem Metadaten schemata mit Eigennamen. Verfügbar unter: <https://www.re3data.org/search?query=%22ABCD+-+Access+to+Biological+Collection+Data%22%7C%22AVM+-+Astronomy+Visualization+Metadata%22%7C%22CF+%28Climate+and+Forecast%29+Metadata+Conventions%22%7C%22CIF+-+Crystallographic+Information+Framework%22%7C%22CIM+-+Common+Information+Model%22%7C%22CSMD-CCLRC+Core+Scientific+Metadata+Model%22%7C%22DCAT+-+Data+Catalog+Vocabulary%22%7C%22DDI+-+Data+Documentation+Initiative%22%7C%22DIF+-+Directory+Interchange+Format%22%7C%22Darwin+Core%22%7C%22DataCite+Metadata+Schema%22%7C%22Dublin+Core%22%7C%22EML+-+Ecological+Metadata+Language%22%7C%22FGDC%2FCSDGM+-+Federal+Geographic+Data+Committee+Content+Standard+for+Digital+Geospatial+Metadata%22%7C%22FITS+-+Flexible+Image+Transport+System%22%7C%22Genome+Metadata%22%7C%22ISA-Tab%22%7C%22ISO+19115%22%7C%22International+Virtual+Observatory+Alliance+Technical+Specifications%22%7C%22MIBBI+-+Minimum+Information+for+Biological+and+Biomedical+Investigations%22%7C%22MIDAS-Heritage%22%7C%22OAI-ORE+-+Open+Archives+Initiative+Object+Reuse+and+Exchange%22%7C%22PROV%22%7C%22QuDEX+-+Qualitative+Data+Exchange+Format%22%7C%22RDF+Data+Cube+Vocabulary%22%7C%22Repository-Developed+Metadata+Schemas%22%7C%22SDMX+-+Statistical+Data+and+Metadata+Exchange%22%7C%22SPASE+Data+Model%22>. [Zugriff am 2024-06-06]. Suchergebnis archiviert unter <https://www.doi.org/10.5281/zenodo.11401021>.

- [34] STARR, Joan und Angela GASTL. isCitedBy: A metadata scheme for DataCite. *D-Lib Magazine*, Bd. 17 (2011), 1/2. ISSN 1082-9873. Verfügbar unter: <https://doi.org/10.1045/january2011-starr>.
- [35] COUSIJN, Helena; Patricia FEENEY; Daniella LOWENBERG; Eleonora PRESANI und Natasha SIMONS. Bringing Citations and Usage Metrics Together to Make Data Count. *Data Science Journal*, Bd. 18 (2019), Nr. 9, S. 1–7. Verfügbar unter: <https://doi.org/10.5334/dsj-2019-009>.
- [36] DATACITE METADATA WORKING GROUP. *DataCite to Dublin Core Mapping 4.4*. DataCite e.V., 2021. Verfügbar unter: <https://doi.org/10.14454/qn00-qx85>.
- [37] HIEMENZ, Bea Maria und Monika KUBEREK. *Evaluation institutioneller Forschungsdaten-Policies in Deutschland: Analyse und Abgleich mit internationalen Empfehlungen*. Technische Universität Berlin, 2018-12. Verfügbar unter: <https://doi.org/10.14279/depositonce-7324.2>.
- [38] BRINEY, Kristin; Abigail GOBEN und Lisa ZILINSKI. Do You Have an Institutional Data Policy?: A Review of the Current Landscape of Library Data Services and Institutional Data Policies. *Journal of Librarianship and Scholarly Communication*, Bd. 3 (), Nr. 2, eP1232. Verfügbar unter: <https://doi.org/10.7710/2162-3309.1232>.
- [39] HIEMENZ, Bea Maria und Monika KUBEREK. Leitlinie? Grundsätze? Policy? Richtlinie? – Forschungsdaten-Policies an deutschen Universitäten. *o-bib. Das offene Bibliotheksjournal*, Bd. 5 (2018), Nr. 2, S. 1–13. Verein Deutscher Bibliothekarinnen und Bibliothekare [VDB]. Verfügbar unter: <https://doi.org/10.5282/o-bib/2018H2S1-13>.
- [40] HOCHSCHULREKTORENKONFERENZ. *Liste der Hochschulen*. Hochschulkompass, 2024. Verfügbar unter: https://hs-kompass.de/kompass/xml/download/hs_liste.txt. [Zugriff am 2024-02-23].
- [41] KRASSNIG, David. *Datensatz zu „Publikationspraktiken für Forschungsdaten in Hochschulschriften: Eine Untersuchung der Veröffentlichungsformate und -methoden“*. Zenodo, 2024-06. Verfügbar unter: <https://doi.org/10.5281/zenodo.11401021>.
- [42] KRASSNIG, David. *Simple CSV Sampler*. Software. Version 1.0.0. GitHub, 2024. Verfügbar unter: <https://github.com/DavidKrassnig/simple-csv-sampler>. [Zugriff am 2024-02-25]. Letzter Commit: 0fae7ec.
- [43] FORSCHUNGSDATEN.ORG. *Data Policies*. 2024. Verfügbar unter: https://www.forschungsdaten.org/index.php?title=Data_Policies&oldid=7865. [Zugriff am 2024-02-27].
- [44] CID, Albert Astals; Carlos Garcia CAMPOS; Adrian JOHNSON u. a. *Poppler*. Software. Version 24.05.0. FreeDesktop.org, 2024. Verfügbar unter: <https://gitlab.freedesktop.org/poppler/poppler>. Letzter Commit: 14191296.

- [45] PEARSON, Karl. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, Bd. 50 (1900), Nr. 302, S. 157–175. Taylor & Francis. Verfügbar unter: <https://doi.org/10.1080/14786440009463897>.
- [46] PEARSON, Karl. On the Theory Of Contingency and Its Relation To Association and Normal Correlation. In: {*Mathematical Contributions to the Theory of Evolution*. 1904.}.
- [47] CRAMÉR, Harald. The two-dimensional case. In: {*Mathematical Methods of Statistics*. Princeton: Princeton University Press, 1946. ISBN 0-691-08004-6.}, S. 260–290.
- [48] R CORE TEAM. *R: A Language and Environment for Statistical Computing*. Software. Version 4.4.0. Wien, Österreich: 2024-04. R Foundation for Statistical Computing. Verfügbar unter: <https://www.R-project.org/>.
- [49] WICKHAM, Hadley; Romain FRANÇOIS; Lionel HENRY; Kirill MÜLLER und Davis VAUGHAN. *dplyr: A Grammar of Data Manipulation*. Software [R-Paket]. Version 1.1.4. [Kein Datum]. Verfügbar unter: <https://github.com/tidyverse/dplyr>. Letzter Commit: 74de244.
- [50] THE DOCUMENT FOUNDATION. *LibreOffice*. Software. Version 24.2.3.2. 2024-05. Verfügbar unter: <https://www.libreoffice.org/>.
- [51] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION [ISO]. *ISO 3166-2:2020: Codes for the representation of names of countries and their subdivisions. Part 2: Country subdivision code*. Standard. Geneva, CH: 2020-08.
- [52] DELLMANN, Sarah. Relevante Faktoren für eine gelungene Implementierung von FDM-Services vor Ort: Ergebnisse einer Befragung von FDM-Mitarbeiter*innen an hessischen Hochschulen. *Bausteine Forschungsdatenmanagement*, 2022-09, Nr. 2. Verfügbar unter: <https://doi.org/10.17192/bfdm.2022.2.8428>. Verfügbar unter: <https://bausteine-fdm.de/article/view/8428>.
- [53] TECHNISCHE INFORMATIONSBIbliothEK [TIB]. *Institutionelle Repositorium der Leibniz Universität Hannover*. 2024. Verfügbar unter: <https://www.repo.uni-hannover.de/oai/request?verb=Identify>. [Zugriff am 2024-03-21].
- [54] KRASSNIG, David. *Simple DSpace5 Sampler*. Software. Version 1.0.0. GitHub, 2024. Verfügbar unter: <https://github.com/DavidKrassnig/simple-dspace5-downloader>. [Zugriff am 2024-05-28]. Letzter Commit: 91b146e.
- [55] PRÄSIDIUM DER GOTTFRIED WILHELM LEIBNIZ UNIVERSITÄT HANNOVER. *Zahlenspiegel 2013 der Leibniz Universität Hannover*. Berichte der Leibniz Universität Hannover, Nr. 13. Gutenberg Beuys Feindruckerei GmbH, 2013. ISSN 1869-1226, 1869-1242. Verfügbar unter: https://www.uni-hannover.de/fileadmin/luh/content/planung_controlling/statistik/zahlenspiegel/zahlenspiegel_2013.pdf. [Zugriff am 2024-05-04].

- [56] PRÄSIDIUM DER GOTTFRIED WILHELM LEIBNIZ UNIVERSITÄT HANNOVER. *Zahlenspiegel 2014 der Leibniz Universität Hannover*. Berichte der Leibniz Universität Hannover, Nr. 14. Gutenberg Beuys Feindruckerei GmbH, 2014. ISSN 1869-1226, 1869-1242. Verfügbar unter: https://www.uni-hannover.de/fileadmin/luh/content/planung_controlling/statistik/zahlenspiegel/zahlenspiegel_2014.pdf. [Zugriff am 2024-05-04].
- [57] PRÄSIDIUM DER GOTTFRIED WILHELM LEIBNIZ UNIVERSITÄT HANNOVER. *Zahlenspiegel 2015 der Leibniz Universität Hannover*. Berichte der Leibniz Universität Hannover, Nr. 15. Gutenberg Beuys Feindruckerei GmbH, 2015. ISSN 1869-1226, 1869-1242. Verfügbar unter: https://www.uni-hannover.de/fileadmin/luh/content/planung_controlling/statistik/zahlenspiegel/zahlenspiegel_2015.pdf. [Zugriff am 2024-05-04].
- [58] PRÄSIDIUM DER GOTTFRIED WILHELM LEIBNIZ UNIVERSITÄT HANNOVER. *Zahlenspiegel 2016 der Leibniz Universität Hannover*. Berichte der Leibniz Universität Hannover, Nr. 16. Gutenberg Beuys Feindruckerei GmbH, 2016. ISSN 1869-1226, 1869-1242. Verfügbar unter: https://www.uni-hannover.de/fileadmin/luh/content/planung_controlling/statistik/zahlenspiegel/zahlenspiegel_2016.pdf. [Zugriff am 2024-05-04].
- [59] PRÄSIDIUM DER GOTTFRIED WILHELM LEIBNIZ UNIVERSITÄT HANNOVER. *Zahlenspiegel 2017 der Leibniz Universität Hannover*. Berichte der Leibniz Universität Hannover, Nr. 17. Gutenberg Beuys Feindruckerei GmbH, 2017. ISSN 1869-1226, 1869-1242. Verfügbar unter: https://www.uni-hannover.de/fileadmin/luh/content/planung_controlling/statistik/zahlenspiegel/zahlenspiegel_2017.pdf. [Zugriff am 2024-05-04].
- [60] PRÄSIDIUM DER GOTTFRIED WILHELM LEIBNIZ UNIVERSITÄT HANNOVER. *Zahlenspiegel 2018 der Leibniz Universität Hannover*. Berichte der Leibniz Universität Hannover, Nr. 18. Gutenberg Beuys Feindruckerei GmbH, 2018. ISSN 1869-1226, 1869-1242. Verfügbar unter: https://www.uni-hannover.de/fileadmin/luh/content/planung_controlling/statistik/zahlenspiegel/zahlenspiegel_2018.pdf. [Zugriff am 2024-05-04].
- [61] PRÄSIDIUM DER GOTTFRIED WILHELM LEIBNIZ UNIVERSITÄT HANNOVER. *Zahlenspiegel 2019 der Leibniz Universität Hannover*. Berichte der Leibniz Universität Hannover, Nr. 19. Gutenberg Beuys Feindruckerei GmbH, 2019. ISSN 1869-1226, 1869-1242. Verfügbar unter: https://www.uni-hannover.de/fileadmin/luh/content/planung_controlling/statistik/zahlenspiegel/zahlenspiegel_2019.pdf. [Zugriff am 2024-05-04].
- [62] PRÄSIDIUM DER GOTTFRIED WILHELM LEIBNIZ UNIVERSITÄT HANNOVER. *Zahlenspiegel 2020 der Leibniz Universität Hannover*. Berichte der Leibniz Universität Hannover, Nr. 20. Gutenberg Beuys Feindruckerei GmbH, 2020. ISSN 1869-1226, 1869-1242. Verfügbar unter: https://www.uni-hannover.de/fileadmin/luh/content/planung_controlling/statistik/zahlenspiegel/zahlenspiegel_2020.pdf. [Zugriff am 2024-05-04].

- [63] PRÄSIDIUM DER GOTTFRIED WILHELM LEIBNIZ UNIVERSITÄT HANNOVER. *Zahlenspiegel 2021 der Leibniz Universität Hannover*. Berichte der Leibniz Universität Hannover, Nr. 21. Gutenberg Beuys Feindruckerei GmbH, 2021. ISSN 1869-1226, 1869-1242. Verfügbar unter: https://www.uni-hannover.de/fileadmin/luh/content/planung_controlling/statistik/zahlenspiegel/zahlenspiegel_2021.pdf. [Zugriff am 2024-05-04].
- [64] PRÄSIDIUM DER GOTTFRIED WILHELM LEIBNIZ UNIVERSITÄT HANNOVER. *Zahlenspiegel 2022 der Leibniz Universität Hannover*. Berichte der Leibniz Universität Hannover, Nr. 22. Gutenberg Beuys Feindruckerei GmbH, 2022. ISSN 1869-1226, 1869-1242. Verfügbar unter: https://www.uni-hannover.de/fileadmin/luh/content/planung_controlling/statistik/zahlenspiegel/zahlenspiegel_2022.pdf. [Zugriff am 2024-05-04].
- [65] PRÄSIDIUM DER GOTTFRIED WILHELM LEIBNIZ UNIVERSITÄT HANNOVER. *Zahlenspiegel 2023 der Leibniz Universität Hannover*. Berichte der Leibniz Universität Hannover, Nr. 23. Gutenberg Beuys Feindruckerei GmbH, 2023. ISSN 1869-1226, 1869-1242. Verfügbar unter: https://www.uni-hannover.de/fileadmin/luh/content/planung_controlling/statistik/zahlenspiegel/luh_zahlenspiegel_2023.pdf. [Zugriff am 2024-05-04].
- [66] SIMUKOVIC, Elena; Raphael THIELE; Alexander STRUCK; Maxi KINDLING und Peter SCHIRMBACHER. *Was sind Ihre Forschungsdaten?* 2014. Verfügbar unter: <https://doi.org/10.18452/13600>.
- [67] KRASSNIG, Sarah Ceylan. Mündliches Gespräch [persönliche Kommunikation]. 2024-04-02.
- [68] ARCHAMBAULT, Éric; Didier AMYOT; Philippe DESCHAMPS; Aurore NICOL; Françoise PROVENCHER u. a. *Proportion of Open Access Papers Published in Peer-Reviewed Journals at the European and World Levels—1996–2013*. European Commission, 2014-10.
- [69] PIWOWAR, Heather; Jason PRIEM; Vincent LARIVIÈRE; Juan Pablo ALPERIN; Lisa MATTHIAS u. a. The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ*, Bd. 6 (2018), e4375. ISSN 2167-8359. Verfügbar unter: <https://doi.org/10.7717/peerj.4375>.
- [70] SEVERIN, Anna; Matthias EGGER; Martin Paul EVE und Daniel HÜRLIMANN. Discipline-specific open access publishing practices and barriers to change: an evidence-based review. Version 2. *F1000Research*, Bd. 7 (2020). Verfügbar unter: <https://doi.org/10.12688/f1000research.17328.2>.
- [71] HAMANN, Hanjo und Daniel HÜRLIMANN. Open Access bei der Veröffentlichung rechtswissenschaftlicher Fachliteratur: Was soll das? In: {HAMANN, Hanjo und Daniel HÜRLIMANN. *Open Access in der Rechtswissenschaft*. 1. Auflage. Baden-Baden: Nomos, 2019. ISBN 978-3-7489-0365-9.}, S. 3–30. Verfügbar unter: <https://doi.org/10.5771/9783748903659-9>.

- [72] FISCHER, Georg. Im Ringen um Erkenntnis und Anerkennung: Wie Rechtswissenschaftler*innen das eigene akademische Publizieren im Zuge von Open Access sehen. Katharina de la DURANTAYE; Ellen EULER; Alexandra KEMMERER; Paul KLIMPEL; Andreas NESTL u. a. *RuZ - Recht und Zugang*, Bd. 3 (2022), Nr. 1, S. 19–49. Baden-Baden: Nomos. ISSN 2699-1284 [Druck], 2942-3368 [Online]. Verfügbar unter: <https://doi.org/10.5771/2699-1284-2022-1-19>.
- [73] EISENTRAUT, Nikolas. Open Access in der Rechtswissenschaft. *Recht, Bibliothek, Dokumentation [RBD]*, Bd. 48 (2018), 2–3, S. 87–92. ISSN 0935-2538. Verfügbar unter: <https://doi.org/10.17169/refubium-26494>.
- [74] STEINHAUER, Eric W. Zur Sichtbarkeit und Verbreitung rechtswissenschaftlicher Dissertationen. In: {HAMANN, Hanjo und Daniel HÜRLIMANN. *Open Access in der Rechtswissenschaft*. 1. Auflage. Baden-Baden: Nomos, 2019. ISBN 978-3-7489-0365-9.}, S. 31–51. Verfügbar unter: <https://doi.org/10.5771/9783748903659-37>.
- [75] BJÖRK, Bo-Christer und David SOLOMON. Open access versus subscription journals: A comparison of scientific impact. *BMC Medicine*, Bd. 10 (2012), Nr. 1, S. 73. ISSN 1741-7015. Verfügbar unter: <https://doi.org/10.1186/1741-7015-10-73>.
- [76] PRÄSIDIUM DER GOTTFRIED WILHELM LEIBNIZ UNIVERSITÄT HANNOVER und SENAT DER GOTTFRIED WILHELM LEIBNIZ UNIVERSITÄT HANNOVER. *Richtlinie zum Umgang mit Forschungsdaten an der Leibniz Universität Hannover*. Version 3.0. 2020-06. Verfügbar unter: https://www.uni-hannover.de/fileadmin/luh/content/dezernat4/FDM/200723_RichtlinieForschungsdatenLUH_v3.pdf. [Zugriff am 2024-03-11].
- [77] LOPEZ, Patrice; Nicolas KIEFFER u. a. *DataSeer-ML*. Software. GitHub, 2023. Verfügbar unter: <https://github.com/DataSeer/dataseer-ml>. [Zugriff am 2024-02-25]. Letzter Commit: 00efd6d.
- [78] SERVICE-TEAM FORSCHUNGSDATEN. *Externe Informations-Portale*. 2023-01. Verfügbar unter: <https://www.fdm.uni-hannover.de/de/materialien/externe-informationen>. [Zugriff am 2024-06-05].

DIESE SEITE WURDE ABSICHTLICH LEER GELASSEN

Appendix

DIESE SEITE WURDE ABSICHTLICH LEER GELASSEN

Anhang A

Richtlinien zu Forschungsdaten aus deutschen Promotionsvorhaben

A.1 Tabellen

Tabelle A.1: Klassifikation der allgemeingültigen Dokumente in relativer Angabe zum Bundesland. Absolute Werte in Klammern angegeben. Angabe der Bundesländer nach ISO 3166-2:2020 [51].

	Keine Verfügbar		GWP-Richtlinien		Andere Richtlinien		FD-Richtlinien	
DE-BW	23,53 %	(4)	41,18 %	(7)	0,00 %	(0)	35,29 %	(6)
DE-BY	16,67 %	(2)	50,00 %	(6)	0,00 %	(0)	33,33 %	(4)
DE-BE	20,00 %	(1)	20,00 %	(1)	0,00 %	(0)	60,00 %	(3)
DE-BB	0,00 %	(0)	25,00 %	(1)	0,00 %	(0)	75,00 %	(3)
DE-HB	12,50 %	(0)	50,00 %	(1)	0,00 %	(0)	50,00 %	(1)
DE-HH	10,00 %	(1)	75,00 %	(6)	0,00 %	(0)	12,50 %	(1)
DE-HE	50,00 %	(1)	40,00 %	(4)	0,00 %	(0)	50,00 %	(5)
DE-MV	9,09 %	(1)	50,00 %	(1)	0,00 %	(0)	0,00 %	(0)
DE-NI	9,09 %	(1)	0,00 %	(0)	0,00 %	(0)	90,91 %	(10)
DE-NW	0,00 %	(2)	40,91 %	(9)	0,00 %	(0)	50,00 %	(11)
DE-RP	0,00 %	(0)	80,00 %	(4)	0,00 %	(0)	20,00 %	(1)
DE-SL	0,00 %	(0)	50,00 %	(1)	0,00 %	(0)	50,00 %	(1)
DE-SN	0,00 %	(0)	60,00 %	(3)	0,00 %	(0)	40,00 %	(2)
DE-ST	0,00 %	(0)	66,67 %	(4)	0,00 %	(0)	33,33 %	(2)
DE-SH	50,00 %	(1)	50,00 %	(1)	0,00 %	(0)	0,00 %	(0)
DE-TH	0,00 %	(0)	50,00 %	(1)	0,00 %	(0)	50,00 %	(1)
Alle	12,17 %	(14)	43,48 %	(50)	0,00 %	(0)	44,35 %	(51)

Tabelle A.2: Klassifikation der promotionsspezifischen Dokumente in relativer Angabe zum Bundesland. Absolute Werte in Klammern angegeben. Angabe der Bundesländer nach ISO 3166-2:2020 [51].

	Kein(e)		GWP-Richtlinie				FD-Richtlinie			
	Zugang	FD-Richtlinie	Empfehlung	Verpflichtung	Empfehlung	Verpflichtung	Empfehlung	Verpflichtung		
DE-BW	0,00 % (0)	5,88 % (1)	0,00 % (0)	82,35 % (14)	0,00 % (0)	11,76 % (2)				
DE-BY	0,00 % (0)	50,00 % (6)	0,00 % (0)	50,00 % (6)	0,00 % (0)	0,00 % (0)				
DE-BE	20,00 % (1)	20,00 % (1)	20,00 % (1)	0,00 % (0)	0,00 % (0)	40,00 % (2)				
DE-BB	0,00 % (0)	25,00 % (1)	0,00 % (0)	50,00 % (2)	0,00 % (0)	25,00 % (1)				
DE-HB	0,00 % (0)	50,00 % (1)	0,00 % (0)	50,00 % (1)	0,00 % (0)	0,00 % (0)				
DE-HH	0,00 % (0)	25,00 % (2)	20,00 % (2)	50,00 % (4)	0,00 % (0)	0,00 % (0)				
DE-HE	10,00 % (1)	20,00 % (2)	0,00 % (0)	20,00 % (2)	10,00 % (1)	40,00 % (4)				
DE-MV	0,00 % (0)	0,00 % (0)	0,00 % (0)	100,00 % (2)	0,00 % (0)	0,00 % (0)				
DE-NI	9,09 % (1)	9,09 % (1)	0,00 % (0)	81,82 % (9)	0,00 % (0)	0,00 % (0)				
DE-NW	0,00 % (0)	40,91 % (9)	0,00 % (0)	50,00 % (11)	0,00 % (0)	9,09 % (2)				
DE-RP	0,00 % (0)	20,00 % (1)	0,00 % (0)	40,00 % (2)	0,00 % (0)	40,00 % (2)				
DE-SL	0,00 % (0)	50,00 % (1)	50,00 % (1)	0,00 % (0)	0,00 % (0)	0,00 % (0)				
DE-SN	0,00 % (0)	20,00 % (1)	0,00 % (0)	60,00 % (3)	0,00 % (0)	20,00 % (1)				
DE-ST	0,00 % (0)	16,67 % (1)	0,00 % (0)	50,00 % (3)	0,00 % (0)	33,33 % (2)				
DE-SH	0,00 % (0)	50,00 % (1)	0,00 % (0)	50,00 % (1)	0,00 % (0)	0,00 % (0)				
DE-TH	50,00 % (1)	0,00 % (0)	0,00 % (0)	50,00 % (1)	0,00 % (0)	0,00 % (0)				
Alle	3,48 % (4)	25,22 % (29)	3,48 % (4)	53,04 % (61)	0,87 % (1)	13,91 % (16)				

Anhang B

Forschungsdaten im Repository der Leibniz Universität Hannover

B.1 Tabellen

Tabelle B.1: FD-Klassifizierung der Dissertationen aus der Stichprobe nach *Sprache* × *Zeitgruppe* aufgegliedert. Angaben relativ zu der Gesamtanzahl der jeweiligen Zeitgruppe. Absolute Werte in Klammern angegeben.

	2012-2015	2016-2019	2020-2023	Alle
Deutsch	66,43 % (277)	61,56 % (269)	40,89 % (240)	54,55 % (786)
Englisch	33,57 % (140)	38,44 % (168)	58,26 % (342)	45,11 % (650)
Andere	0,00 % (0)	0,00 % (0)	0,85 % (5)	0,35 % (5)

Tabelle B.2: FD-Klassifizierung der Dissertationen aus der Stichprobe nach *Fakultät* × *Klassifikationsstufe* aufgegliedert. Angabe relativ zu der respektiven Gesamtanzahl für *Fakultät*. Absolute Werte in Klammern angegeben.

	Stufe 1	Stufe 2	Stufe 3	Keine
Fakultät ARC	13,33 % (8)	0,00 % (0)	46,67 % (28)	40,00 % (14)
Fakultät BAU	25,45 % (28)	11,82 % (13)	11,82 % (13)	50,91 % (51)
Fakultät INF	28,97 % (42)	12,41 % (18)	17,93 % (26)	40,69 % (54)
Fakultät JUR	0,00 % (0)	0,00 % (0)	0,00 % (0)	100,00 % (0)
Fakultät MAS	12,40 % (15)	4,13 % (5)	19,01 % (23)	64,46 % (77)
Fakultät MAT	13,52 % (33)	4,10 % (10)	20,08 % (49)	62,30 % (122)
Fakultät NAT	22,17 % (102)	20,87 % (96)	40,43 % (186)	16,52 % (71)
Fakultät PHI	12,90 % (16)	0,00 % (0)	19,35 % (24)	67,74 % (31)
Fakultät WIWI	4,62 % (8)	0,58 % (1)	13,87 % (24)	80,92 % (63)
Alle	17,49 % (252)	9,92 % (143)	25,88 % (373)	46,70 % (483)

Tabelle B.3: FD-Klassifizierung der Dissertationen aus der Stichprobe nach *Publikationsart* × *Klassifikationsstufe* × *Jahresgruppe* aufgegliedert. Angaben relativ zu der Gesamtanzahl der Jahresgruppe. Absolute Werte in Klammern angeben.

		2012-2015		2016-2019		2020-2023		Alle	
Intern	Stufe 1	11,99 %	(50)	12,36 %	(54)	9,37 %	(55)	11,03 %	(159)
	Stufe 2	11,51 %	(48)	12,36 %	(54)	11,58 %	(68)	11,80 %	(170)
	Stufe 3	31,89 %	(133)	29,06 %	(127)	21,29 %	(125)	26,72 %	(385)
	Keine	44,60 %	(186)	46,22 %	(202)	57,75 %	(339)	50,45 %	(727)
Beilage	Stufe 1	1,20 %	(5)	2,29 %	(10)	1,19 %	(7)	1,53 %	(22)
	Stufe 2	0,00 %	(0)	0,00 %	(0)	0,00 %	(0)	0,00 %	(0)
	Stufe 3	0,48 %	(2)	0,23 %	(1)	0,51 %	(3)	0,42 %	(6)
	Keine	98,32 %	(410)	97,48 %	(426)	98,30 %	(577)	98,06 %	(1413)
Extern	Stufe 1	0,96 %	(4)	2,75 %	(12)	12,61 %	(74)	6,25 %	(90)
	Stufe 2	0,00 %	(0)	0,00 %	(0)	0,17 %	(1)	0,07 %	(1)
	Stufe 3	0,00 %	(0)	0,23 %	(1)	0,17 %	(1)	0,14 %	(2)
	Keine	99,04 %	(413)	97,03 %	(424)	87,05 %	(511)	93,55 %	(1348)
Alle	Stufe 1	13,19 %	(55)	16,93 %	(74)	20,95 %	(123)	17,49 %	(252)
	Stufe 2	11,51 %	(48)	10,53 %	(46)	8,35 %	(49)	9,92 %	(143)
	Stufe 3	31,65 %	(132)	28,60 %	(125)	19,76 %	(116)	25,88 %	(373)
	Keine	43,65 %	(182)	43,94 %	(192)	50,94 %	(299)	46,70 %	(673)

B.2 Grafiken

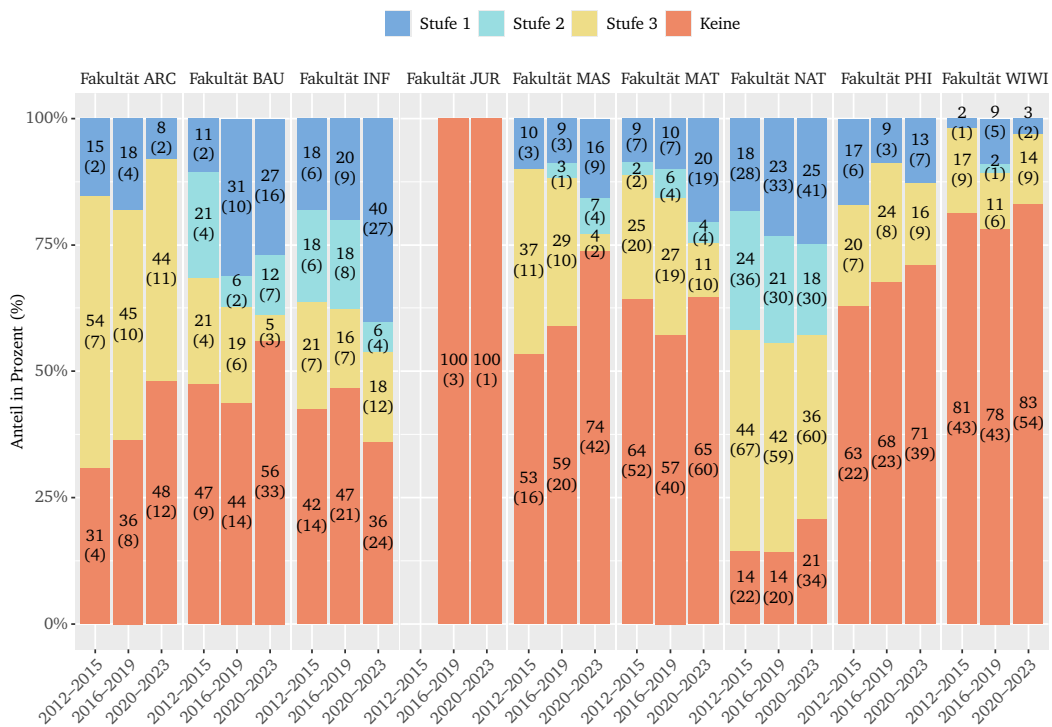


Abbildung B.1: FD-Klassifikation der Dissertationen aus der Stichprobe nach Fakultät, Zeitgruppe und Klassifikationsstufe. Die Höhe der Barren entsprechen dem relativen Anteil zur jeweiligen Fakultät × Zeitgruppe-Gesamtanzahl. Absolute Werte in Klammern angegeben.

DIESE SEITE WURDE ABSICHTLICH LEER GELASSEN



Name: Krassnig Vorname: David

Matr.Nr.: 632351

Eidesstattliche Erklärung zur

- Hausarbeit ***
- Bachelorarbeit ***
- Masterarbeit ***
- Abschlussarbeit im Bibliotheksreferendariat ***

* Die eingereichte PDF-Datei ist mit den Printexemplaren identisch.

Ich erkläre ausdrücklich, dass es sich bei der von mir eingereichten schriftlichen Arbeit mit dem Titel

..... Publikationspraktiken für Forschungsdaten in Hochschulschriften: Eine Untersuchung der
..... Veröffentlichungsformate und -methoden

.....
um eine von mir erstmalig, selbstständig und ohne fremde Hilfe verfasste Arbeit handelt.

Ich erkläre ausdrücklich, dass ich *sämtliche* in der oben genannten Arbeit verwendeten fremden Quellen, auch aus dem Internet (einschließlich Tabellen, Grafiken u. Ä.) als solche kenntlich gemacht habe. Insbesondere bestätige ich, dass ich ausnahmslos sowohl bei wörtlich übernommenen Aussagen bzw. unverändert übernommenen Tabellen, Grafiken u. Ä. (Zitaten) als auch bei in eigenen Worten wiedergegebenen Aussagen bzw. von mir abgewandelten Tabellen, Grafiken u. Ä. anderer Autorinnen und Autoren (Paraphrasen) die Quelle angegeben habe.

Mir ist bewusst, dass Verstöße gegen die Grundsätze der Selbstständigkeit als Täuschung betrachtet und entsprechend der Prüfungsordnung und/oder der Fächerübergreifenden Satzung zur Regelung von Zulassung, Studium und Prüfung (ZSP-HU) geahndet werden.

Datum 12.06.2024

Unterschrift 