


e·infra
austria

AP 5 Metadaten

Umgang mit Metadaten in Repositorien – eine österreichweite Umfrage Zweite Folge

Susanne Blumesberger 
Alexander Zartl

Universität Wien

September 2019



Der Inhalt dieser Veröffentlichung steht unter einer Creative Commons
Namensnennung 4.0 Lizenz (<http://creativecommons.org/licenses/by/4.0/>)

[DOI 10.25651/1.2019.0018](https://doi.org/10.25651/1.2019.0018)

Über e-Infrastructures Austria Plus

Das Projekt „e-Infrastructures Austria Plus“ (2017-2019) ist ein vom österreichischen Bundesministerium für Bildung, Wissenschaft und Forschung (<https://bildung.bmbwf.gv.at/>) gefördertes Projekt von neun österreichischen Universitäten. Ziel des Projekts ist der koordinierte Aufbau eines österreichischen Netzwerks zur Einrichtung und Weiterentwicklung gemeinsamer e-Infrastrukturen durch Bündelung von Ressourcen und vorhandenem Wissen.

Inhalt

Ziel	4
Vorbemerkung	4
Liste der befragten Institutionen.....	5
Ergebnisse	6
1.Allgemeine Fragen.....	6
1.1.Welche Arten von Daten werden an Ihrer Einrichtung gesammelt?.....	6
1.2.Gibt es bereits ein gemeinsames Archiv bzw. Repositorium für die Angehörigen Ihrer Einrichtung? Wenn ja, um welches System handelt es sich?.....	7
1.3.Sind Sie mit der Wahl zufrieden? Wenn nicht: was fehlt?.....	8
1.4.Gibt es fachspezifische Repositorien an Ihrer Einrichtung? Wenn ja, für welche Fächer?	9
1.5.Für wie lange sollen digitale Daten aufbewahrt werden? Können UserInnen die Daten selbst löschen?	10
1.6.Zu welchem Zweck werden die Daten vorrangig archiviert? (Re-Use der Daten / Vorgabe der FördergeberInnen / eigene Forschung / Informationsaustausch / Publikation / andere)	11
1.7.Gibt es eine Schnittstelle zwischen Bibliothekskatalog und Repositorium bzw. ist eine solche in Planung?	12
1.8.Hat die DSGVO konkrete Auswirkungen auf den Umgang mit Repositorien? Z.B.: Gibt es Daten, für die nach einer bestimmten Zeit eine Löschverpflichtung besteht? Wie gehen Sie mit personenbezogenen Daten um?	13
1.9.Gibt es in Ihrer Einrichtung eine Forschungsdaten-Policy?	14
1.10.Werden an Ihrer Einrichtung permanente Identifier (DOIs) vergeben bzw. sind solche in Planung? Wenn ja, welche?	15
1.11.Bieten Sie in Ihrer Einrichtung Beratung im Bereich Datenmanagement an?.....	16
1.12.Gibt es ethische Vorgaben bezüglich des Umgangs mit sensiblen Daten? Wenn ja, welche? (Beispiele: die Möglichkeit zum Verbergen der Metadaten, Embargo, Safe Room).....	17
2.Fragen mit Bezug auf Metadaten.....	18
2.1.Gibt es für die Metadaten eine spezifische Policy?.....	18
2.2.Werden für Metadaten Creative-Commons-Lizenzen oder andere (z.B. GNU) vergeben?	19
2.3.Gibt es eine vollständige öffentlich zugängliche Dokumentation bzw. Beschreibung (Guidelines) der Metadaten? Wenn nein, ist sie in Planung?	20
2.4.Welche Metadatenstandards und Schemata werden innerhalb Ihrer Institution für Ihr Repositorium verwendet, bzw. welche planen Sie?	21
2.5.Bieten Sie Schnittstellen für den Austausch der Metadaten an? Wenn ja, welche?	22
2.6.Wer trägt die Metadaten in das System ein?	24

2.7. Wenn Metadaten direkt aus anderen Systemen in das Repository übernommen werden: mit welcher Schnittstelle?	25
2.8. Gibt es die Möglichkeit zur Verwendung von Templates bzw. Vorlagen als Ausfüllhilfe?	26
2.9. Gibt es für die Metadaten eine formale und / oder inhaltliche Qualitätskontrolle?.....	27
2.10. Verwenden Sie Thesauri oder ein sonstiges kontrolliertes Vokabular bzw. eine freie Beschlagwortung? Wenn ja, welche?.....	28
2.11. Werden zurzeit auch Daten ohne Metadaten gespeichert?.....	29
2.12. Sind die Metadaten barrierefrei nutzbar? Wenn ja, wie wurde das erreicht? Wenn ja, könnten Sie uns bitte ein Beispiel schicken?.....	30
Fazit	31

Ziel des Arbeitspakets „Metadaten“

Ziel des Arbeitspakets „Metadaten“ ist die Erarbeitung eines Metadatenframeworks für Repositorien in Österreich und von Guidelines, die mit FAIR-Richtlinien (FAIR Guiding Principles for scientific data management and stewardship) arbeiten. Die Suche von Forschungsdaten und Projektkooperationen können repositorienübergreifend nur funktionieren, wenn ein gemeinsamer Satz von Metadaten beim Einbringen der Daten erstellt wird und wenn die eingebrachten Daten mit Kontrolliertem Vokabular beschrieben werden. Das Projekt will diese beiden Voraussetzungen schaffen.

Vorbemerkung

Nachdem eine Umfrage zum Thema „Umgang mit Metadaten in Repositorien an österreichischen Universitäten“ schon Bestandteil der ersten Auflage von e-Infrastructures Austria war¹, sollte die Fragestellung im Rahmen der Fortsetzung des Projekts (unter dem Titel e-Infrastructures Austria Plus) erneut behandelt werden. Unter Berücksichtigung des Zeitraumes von rund drei Jahren, der plangemäß zwischen den beiden Befragungen liegen würde, entstand der Gedanke, die stattgefundene Entwicklung zu dokumentieren und die Fortschritte im Vergleich der beiden Fragebogenpakete herauszuarbeiten. Um dies zu ermöglichen, wurde entschieden, auch jene Institutionen zur Befragung einzuladen, die zwar beim ersten Projektteil, aber nicht mehr bei e-Infrastructures Austria Plus dabei waren.

Als weitere Maßnahme zur Wahrung der Vergleichbarkeit wurde der Fragebogen nicht komplett neu gestaltet, sondern im Wesentlichen nur leicht überarbeitet, wobei einige Fragen – basierend auf den Erfahrungen des Jahres 2015 – ergänzt bzw. weggelassen wurden. Auch die Form des Telefoninterviews und die sonstige Vorgangsweise wurden beibehalten, insbesondere die Möglichkeit, den Fragebogen schon vorab unter Hinzuziehung von ExpertInnen an der eigenen Institution auszufüllen und eventuelle Unklarheiten im Zuge des persönlichen Gesprächs zu klären. Ebenso wurde den InterviewpartnerInnen wieder Anonymität zugesichert, d.h. die Rohdaten aus den einzelnen Fragebögen werden archiviert, aber nicht ungefiltert an die Öffentlichkeit weitergegeben.

Der adaptierte Fragebogen wurde Anfang August 2018 ausgesickt, erste Interviews fanden im September statt; abgesehen von einigen wenigen NachzüglerInnen konnte die Aktion kurz vor Weihnachten abgeschlossen werden.

Ohne an dieser Stelle schon auf Details einzugehen, kann gesagt werden, dass die österreichische Universitätslandschaft während der drei Jahre von 2015 bis 2018 bezüglich des Themas „digitale Repositorien“ einen großen Schritt nach vorne gemacht hat. Alle befragten Einrichtungen verfügen mittlerweile über ein solches – was 2015 noch nicht der Fall war – und der Bereich der Forschungsdaten ist vom Randbereich ins Zentrum der Aufmerksamkeit gerückt.

Dem Vergleich 2015 – 2018 wird in jedem Abschnitt jeweils ein eigener Absatz gewidmet. LeserInnen, die sich dafür interessieren, wird empfohlen, ergänzend die zu den Fragen gehörigen Anmerkungen im ersten Bericht (vgl. die untenstehende Fußnote) zu lesen.

Eine Anmerkung zu den gewählten Diagrammtypen: Balkendiagramme kommen stets dort zum Einsatz, wo Mehrfachantworten möglich waren und die Summe der Prozentzahlen daher größer als 100 sein kann. Entscheidungsfragen mit nur einer Antwortmöglichkeit werden durch Kreisdiagramme veranschaulicht.

Wir bedanken uns bei allen GesprächspartnerInnen für die freundliche Unterstützung des Projekts.

¹ Susanne Blumesberger, Alexander Zartl: Umgang mit Metadaten in Repositorien – Eine österreichweite Umfrage. Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare, Bd. 70 (2017), Nr. 2. <https://doi.org/10.31263/voebm.v70i2>

Liste der befragten Institutionen

Im Gegensatz zum ersten Teil des Projekts, bei dem alle teilnehmenden Einrichtungen für die Befragung gewonnen werden konnten, wurden in der zweiten Runde nur 10 von 15 ProjektpartnerInnen erreicht. Andererseits erklärten auch einige Institutionen, die bei der Neuauflage des Projekts gar nicht mehr dabei waren, ihre Bereitschaft zur Mitarbeit, sodass sich folgende Teilnehmerliste ergibt:

1. Akademie der Bildenden Künste
2. Arbeiterkammer Wien
3. Institute of Science and Technology Austria (IST)
4. Medizinische Universität Wien
5. Österreichische Akademie der Wissenschaften
6. Österreichische Nationalbibliothek
7. Technische Universität Graz
8. Technische Universität Wien
9. Universität für angewandte Kunst Wien
10. Universität für Bodenkultur Wien
11. Universität für künstlerische und industrielle Gestaltung Linz
12. Universität für Musik und darstellende Kunst Graz
13. Universität Graz
14. Universität Innsbruck
15. Universität Linz
16. Universität Mozarteum Salzburg
17. Universität Salzburg
18. Universität Wien
19. Wirtschaftsuniversität Wien
20. Medizinische Universität Innsbruck

Im Hinblick auf die Kontinuität der Umfrage kann festgehalten werden, dass 19 von 24 Einrichtungen des ersten Projektteils und eine neu hinzugekommene Einrichtung (die Medizinische Universität Innsbruck) daran teilgenommen haben.

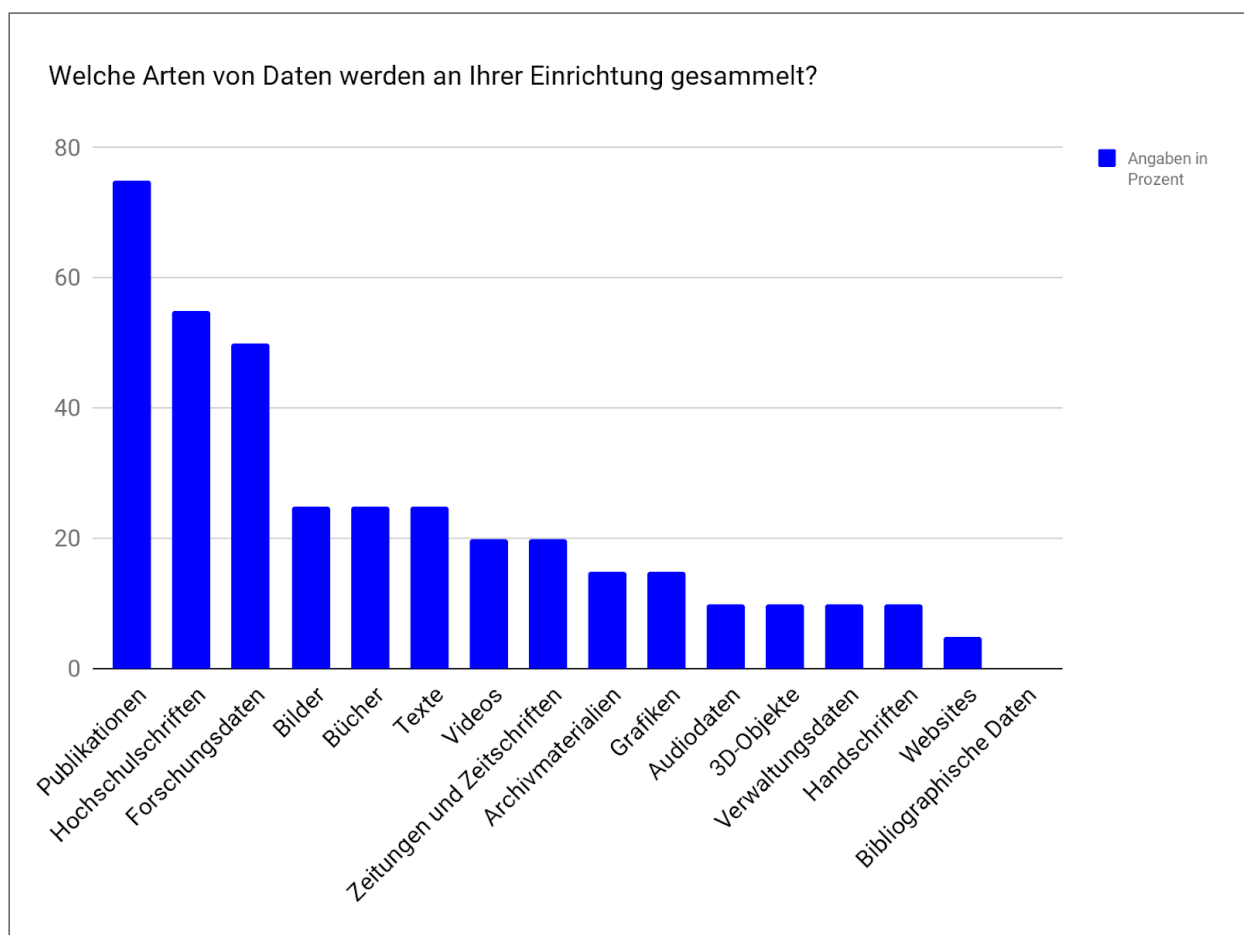
Ergebnisse

1. Allgemeine Fragen

1.1. Welche Arten von Daten werden an Ihrer Einrichtung gesammelt?

Bei dieser Frage gab es kein vorgegebenes Antwortvokabular, sondern es wurden die von den InterviewpartnerInnen gewählten Bezeichnungen übernommen. Daraus ergibt sich eine unscharfe Kategorienbildung mit Überlappungen zwischen den Begriffen „Bilder“ und „Grafiken“ oder „Texte“ und „Publikationen“ etc. Diese Vorgangsweise wurde gewählt, um zu verhindern, dass bestimmte Arten von Daten von vornherein ausgeschlossen werden.

Vergleich: Es zeigt sich ein starker Zuwachs bei den institutionellen Publikationen (von 42% auf 75%) sowie bei den Forschungsdaten (21% auf 50%), die nun gemeinsam mit den Hochschulschriften die Trias der wichtigsten Sammlungsobjekte bilden.

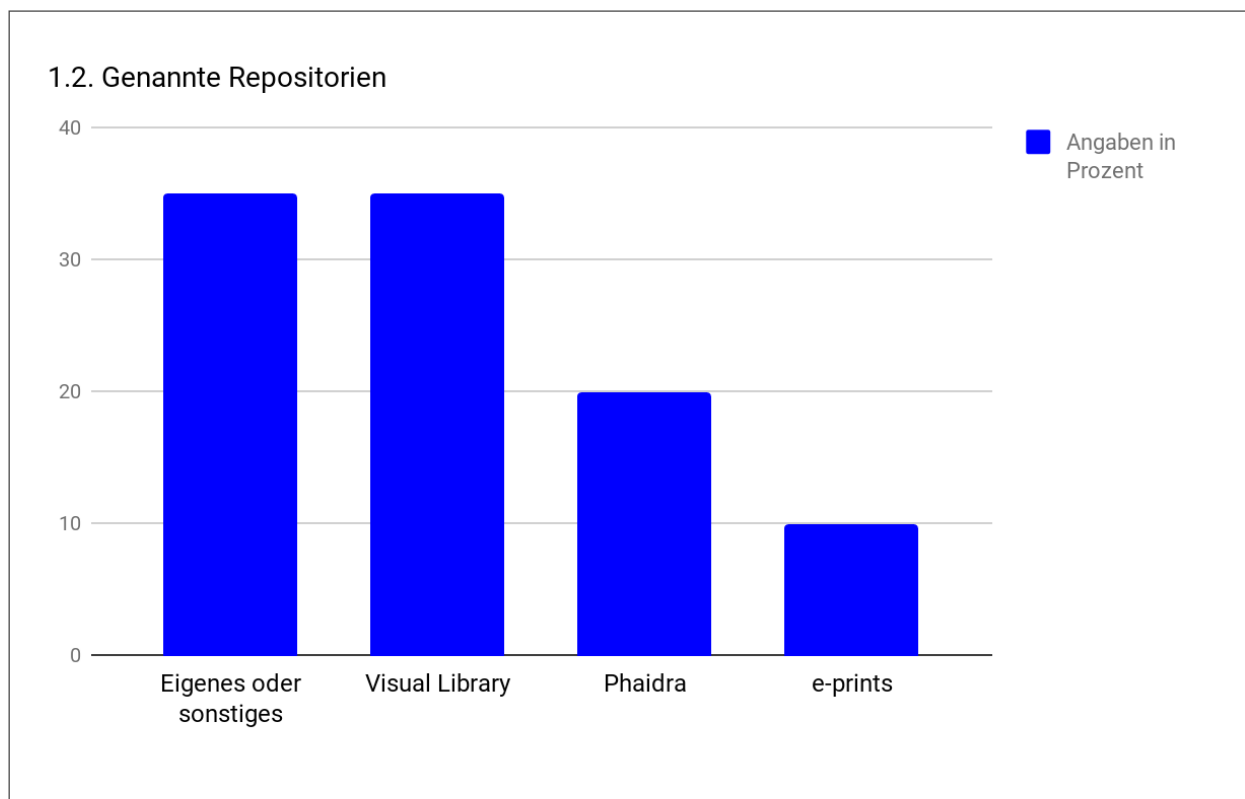


1.2. Gibt es bereits ein gemeinsames Archiv bzw. Repositorium für die Angehörigen Ihrer Einrichtung? Wenn ja, um welches System handelt es sich?

100% der befragten Einrichtungen verfügen über ein (oder mehrere) Repositorien, weshalb das zugehörige Diagramm hier weggelassen werden kann. Das Sammeln digitaler Objekte ist nunmehr Standard. Visual Library und Phaidra dominieren, weitere Systeme kamen höchstens auf zwei (e-prints) oder gar nur eine Nennung (e-pub, Pure, goobi..., subsumiert unter „Eigenes oder sonstige“).

Gezählt wurde auch die (einmal genannte) Nutzung des von CERN gehosteten Systems Zenodo², da dieses ein vollwertiges, mit anderen Produkten durchaus vergleichbares Repositorium darstellt.

Vergleich: Keine wesentlichen Änderungen bei den genutzten Systemen.

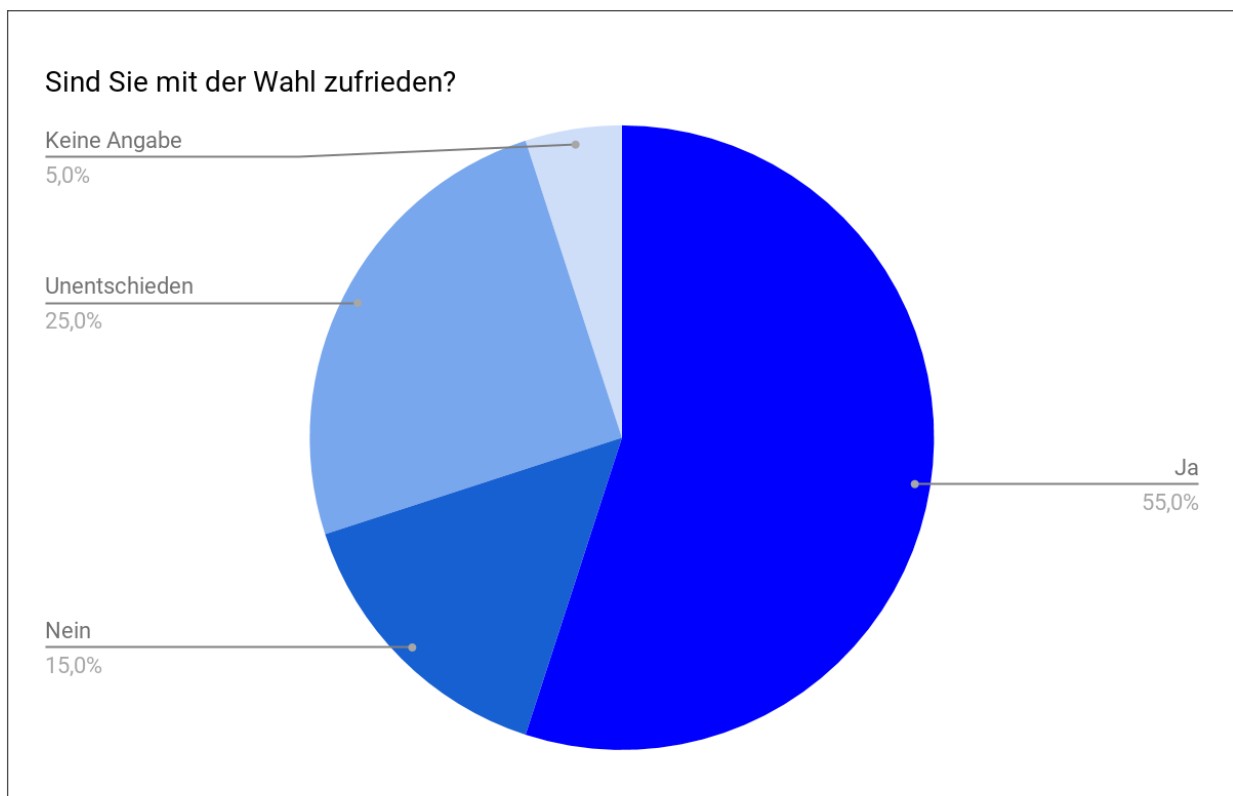


²Siehe <https://zenodo.org/>

1.3. Sind Sie mit der Wahl zufrieden? Wenn nicht: was fehlt?

Mehr als die **Hälfte** der Einrichtungen ist – ggf. trotz leichter Mängel – mit ihren Systemen zufrieden, nur 15% haben überwiegend schlechte Erfahrungen gemacht. In einem Fall gab es bereits einen Systemwechsel. Unter den genannten Mängeln waren Probleme mit dem Herstellersupport (4-mal); unzureichende Ausstattung mit Metadatenformaten (3-mal); unzureichende Ausstattung mit Dateiformaten (2-mal); praxisfremdes Rechtemanagement; fehlende Datenschnittstellen; schlechte Durchsuchbarkeit; zu hohe Herstellerabhängigkeit; fehlender Bulkupload (je einmal).

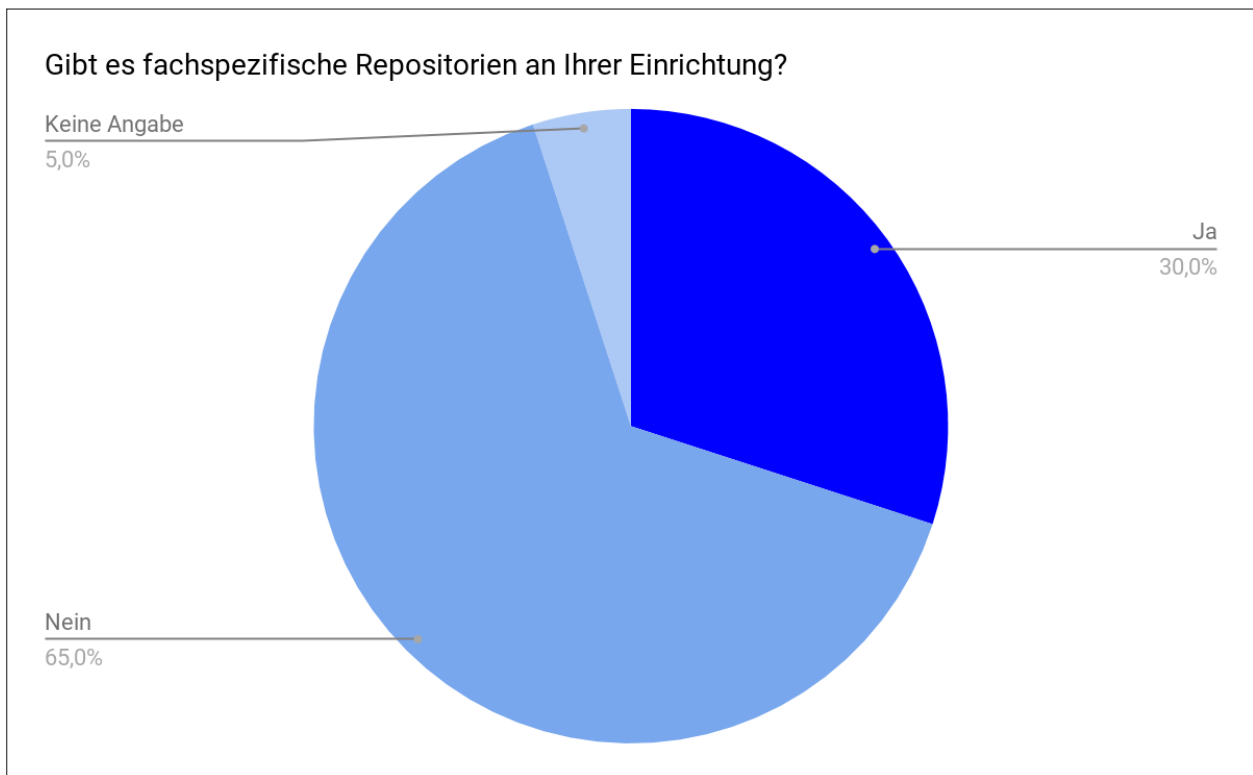
Vergleich: nicht möglich, da diese Frage im ersten Projektteil nicht gestellt wurde.



1.4. Gibt es fachspezifische Repositorien an Ihrer Einrichtung? Wenn ja, für welche Fächer?

Hier wurden auch jene Einrichtungen gezählt, die solche Repositorien nicht selbst betreiben, sondern in Kooperation mit oder als Gast von Partnerinstitutionen. Ein Drittel der Einrichtungen macht von dieser Möglichkeit Gebrauch. Bezüglich der Fachrichtungen wurden genannt: Digital humanities (2-mal), Erdwissenschaften, Filmarchiv, Kunstsammlung, Gewässerökologie und Klimaforschung.

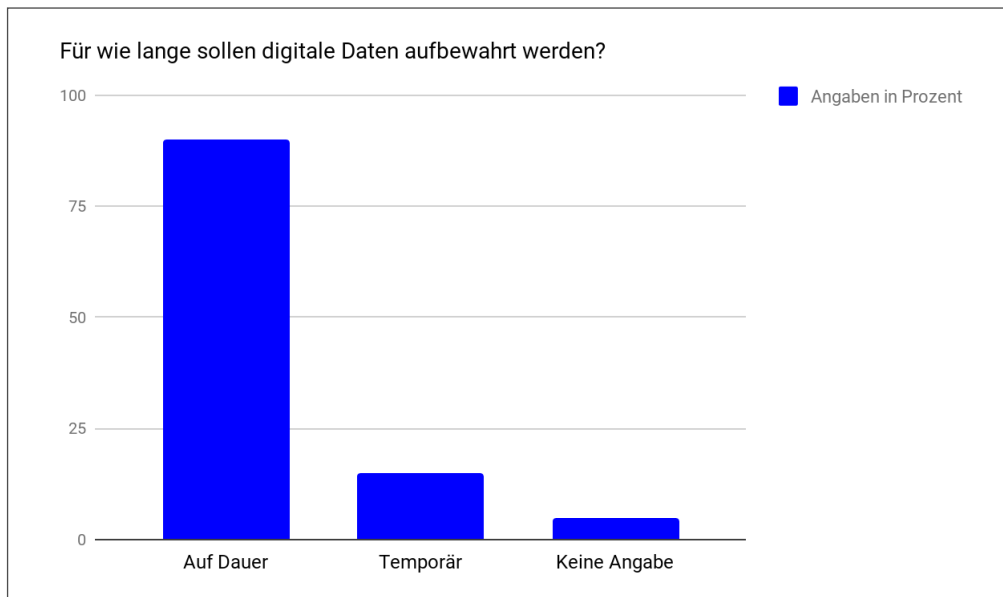
Vergleich: nicht möglich, da diese Frage im ersten Projektteil nicht gestellt wurde.



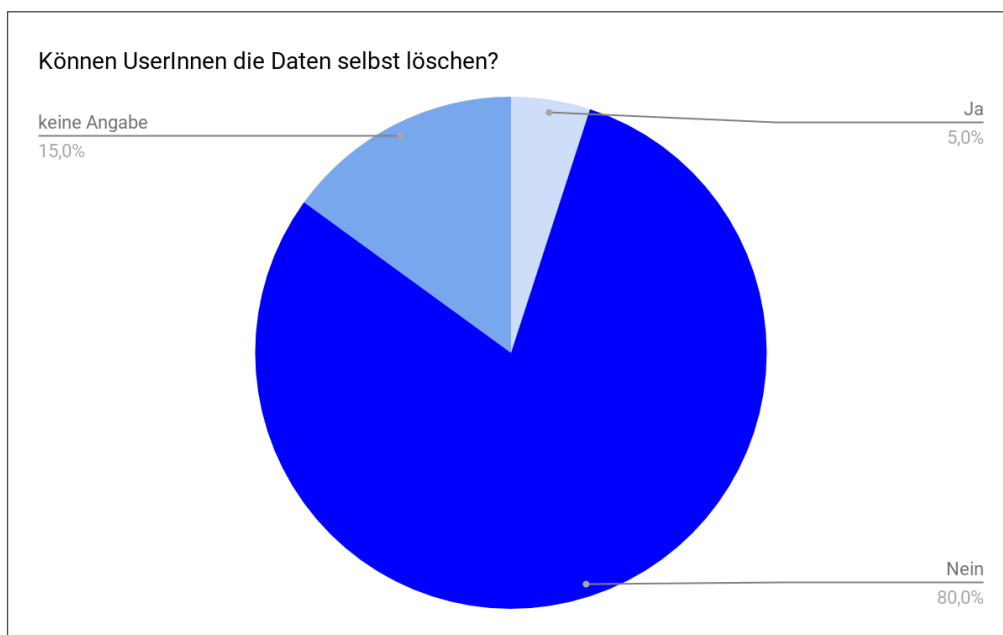
1.5. Für wie lange sollen digitale Daten aufbewahrt werden? Können UserInnen die Daten selbst löschen?

Langzeitarchivierung ist das vorherrschende Paradigma an der überwiegenden Mehrheit der Einrichtungen – nur in einem Fall ist dies explizit nicht der Zweck des eingesetzten Repositoriums. Die Möglichkeit zur vorübergehenden Speicherung von Arbeitsdokumenten, Rohfassungen und sonstigen Zwischenergebnissen mit Ablaufdatum ist nur in zwei Fällen gegeben, der Wunsch danach wurde allerdings öfter angesprochen.

Vergleich: So gut wie keine Änderung feststellbar.



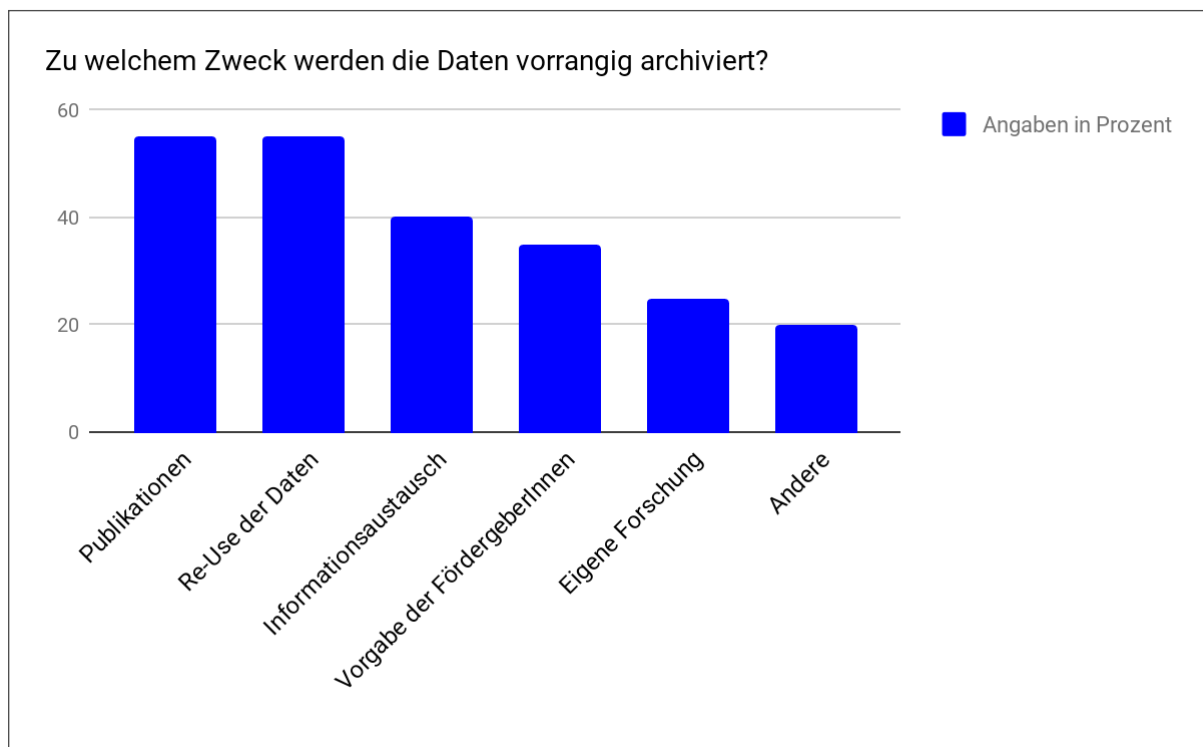
Nur eine Einrichtung bietet den BenutzerInnen des Repositoriums die Möglichkeit zu Selbstlöschung. Löschungen auf Verlangen werden in den meisten Fällen durchgeführt, doch gemäß dem Zweck der Langzeitarchivierung nur unter Angabe von triftigen Gründen.



1.6. Zu welchem Zweck werden die Daten vorrangig archiviert? (Re-Use der Daten / Vorgabe der FördergeberInnen / eigene Forschung / Informationsaustausch / Publikation / andere)

Die Darstellung der Publikationstätigkeit der jeweiligen Einrichtung ist wenig überraschend eines der Hauptziele der digitalen Archivtätigkeit. Ex aequo genannt wurde aber auch die Wiederverwendung (Re-Use) der Daten, was möglicherweise als Hinweis auf das Erstarren des Sharing-Gedankens (Open Science) gesehen werden kann.

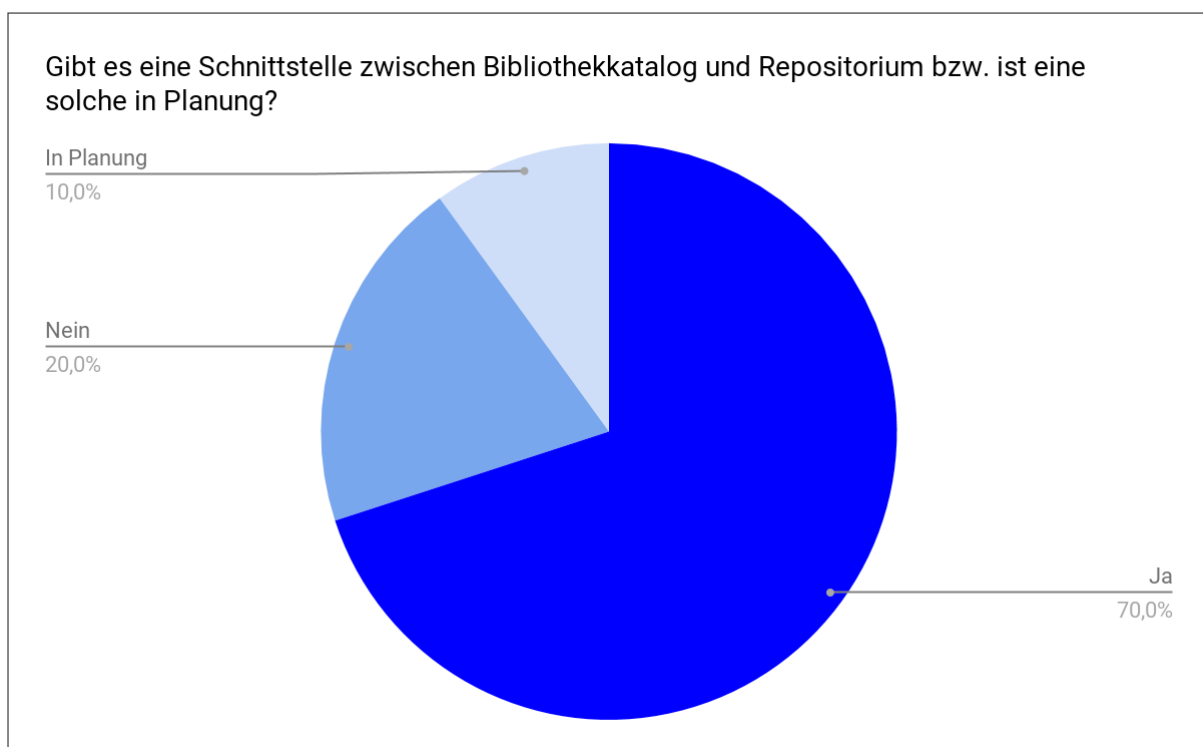
Vergleich: nur eingeschränkt möglich, da diese Frage im ersten Projektteil anders formuliert war, nämlich mit einer anderen Vorgabe der möglichen Zwecke. Die Antworten „Herausgabe von Publikationen“ und „Informationsaustausch“ erzielten 2015 ähnliche Werte, wohingegen der Zweck „Eigene Forschungsziele“ von 79% auf 25% absackte.



1.7. Gibt es eine Schnittstelle zwischen Bibliothekskatalog und Repository bzw. ist eine solche in Planung?

Vier Fünftel der befragten Einrichtungen verfügen über eine Schnittstelle zwischen Repository und Bibliothekskatalog oder haben eine solche in Planung.

Vergleich: Hier ist überraschender Weise – trotz des immer noch hohen Wertes – eine Rückentwicklung festzustellen: Während bei der ersten Umfrage noch 96% der Befragten angaben, eine Schnittstelle zu betreiben oder zu planen, sind es drei Jahre später nur mehr 80%. In zwei Fällen ist dies darauf zurückzuführen, dass die Planung (z.B. nach einem Systemwechsel) offenbar aufgegeben wurde³; ein weiteres „nein“ kommt von einer Einrichtung, die 2015 noch nicht an der Befragung teilgenommen hat.



³ Der Wechsel von ALEPH zu Alma reicht für sich genommen als Erklärung nicht aus, da auch in Alma die Implementierung einer solchen Schnittstelle möglich ist.

1.8. Hat die DSGVO konkrete Auswirkungen auf den Umgang mit Repositorien?
Z.B.: Gibt es Daten, für die nach einer bestimmten Zeit eine Löschverpflichtung besteht? Wie gehen Sie mit personenbezogenen Daten um?

Zwei Drittel der Einrichtung erklären, keine Auswirkungen der DSGVO auf ihre Repositoriumsaktivitäten zu sehen. Dies wird dadurch erklärt, dass personenbezogene Daten in den Repositorien gar nicht erst erfasst werden oder – dort, wo dies doch der Fall ist, z.B. bei den Forschungs- und PatientInnen Daten an medizinischen Universitäten – ein sehr hohes Schutzniveau schon bisher verpflichtend war. Allgemein üblich sind überdies folgende Maßnahmen: Löschung persönlicher Daten auf Antrag der Betroffenen; Löschung nach Ende eines Arbeitsverhältnisses; Löschung von nicht mehr benötigten Adress- u. Kontaktdaten.

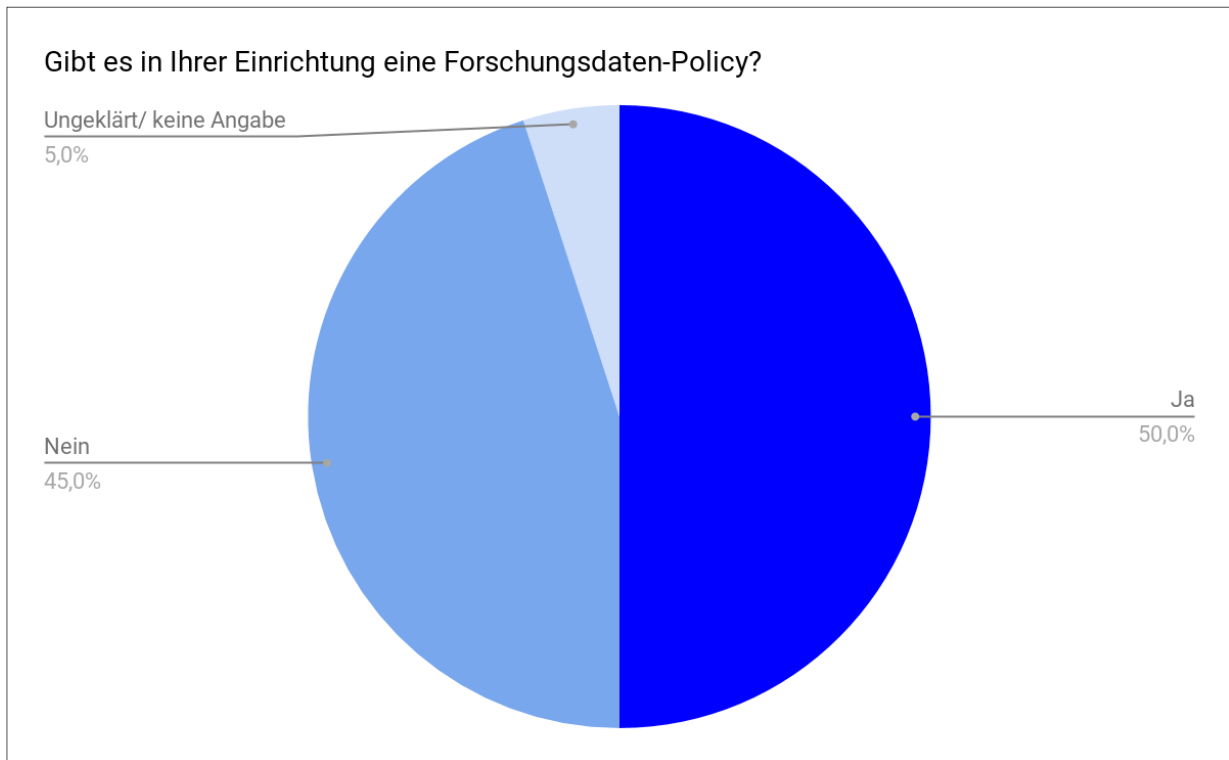
Vergleich: nicht möglich, da diese Frage im ersten Projektteil nicht gestellt wurde.



1.9. Gibt es in Ihrer Einrichtung eine Forschungsdaten-Policy?

Als „Ja“ gewertet wurden nur jene Antworten, wo eine Forschungsdaten-Policy bereits beschlossen und in Kraft ist; die bloße Planung wurde als „Nein“ gezählt. Immerhin die Hälfte der Einrichtungen erfüllt dieses Kriterium.

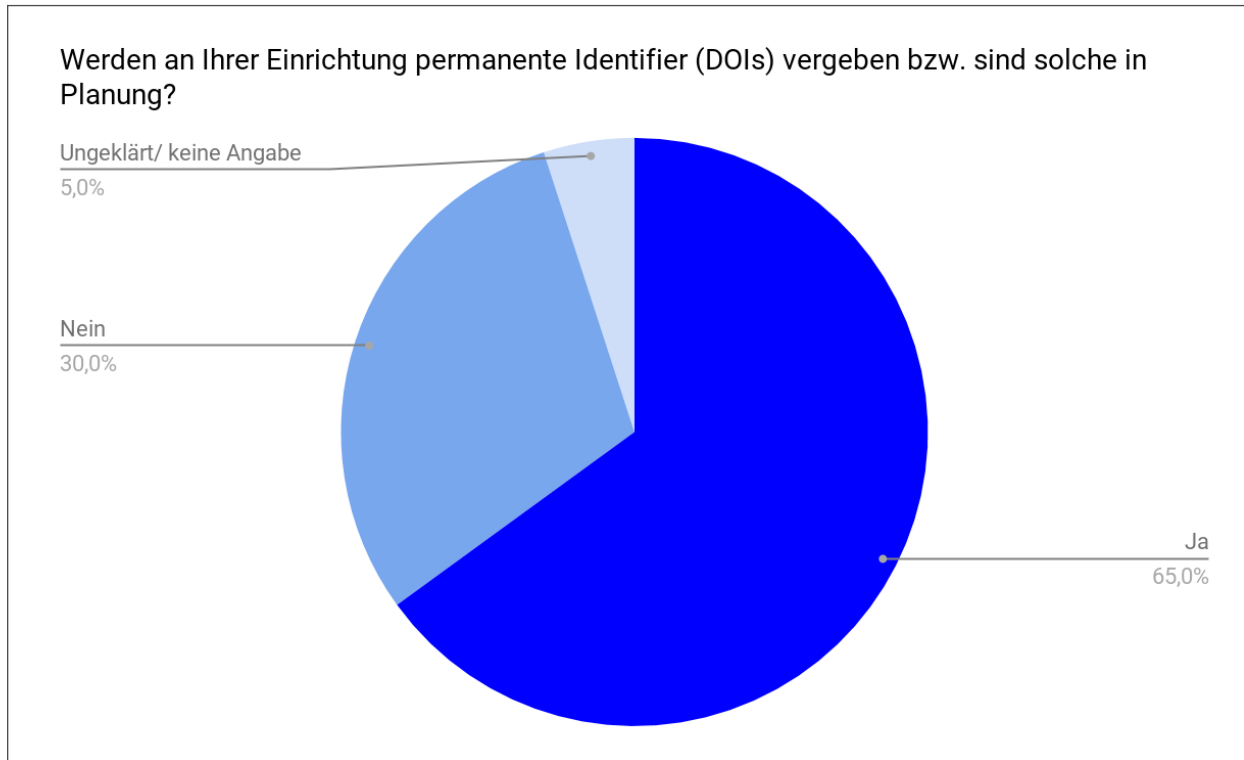
Vergleich: nicht möglich, da diese Frage im ersten Projektteil nicht gestellt wurde.



1.10. Werden an Ihrer Einrichtung permanente Identifier (DOIs) vergeben bzw. sind solche in Planung? Wenn ja, welche?

Zwei Drittel der Einrichtungen vergeben Handles in irgendeiner Form. Genannt wurden DOIs und URNs (je 8-mal); drei Einrichtungen vergeben sowohl DOIs als auch URNs.

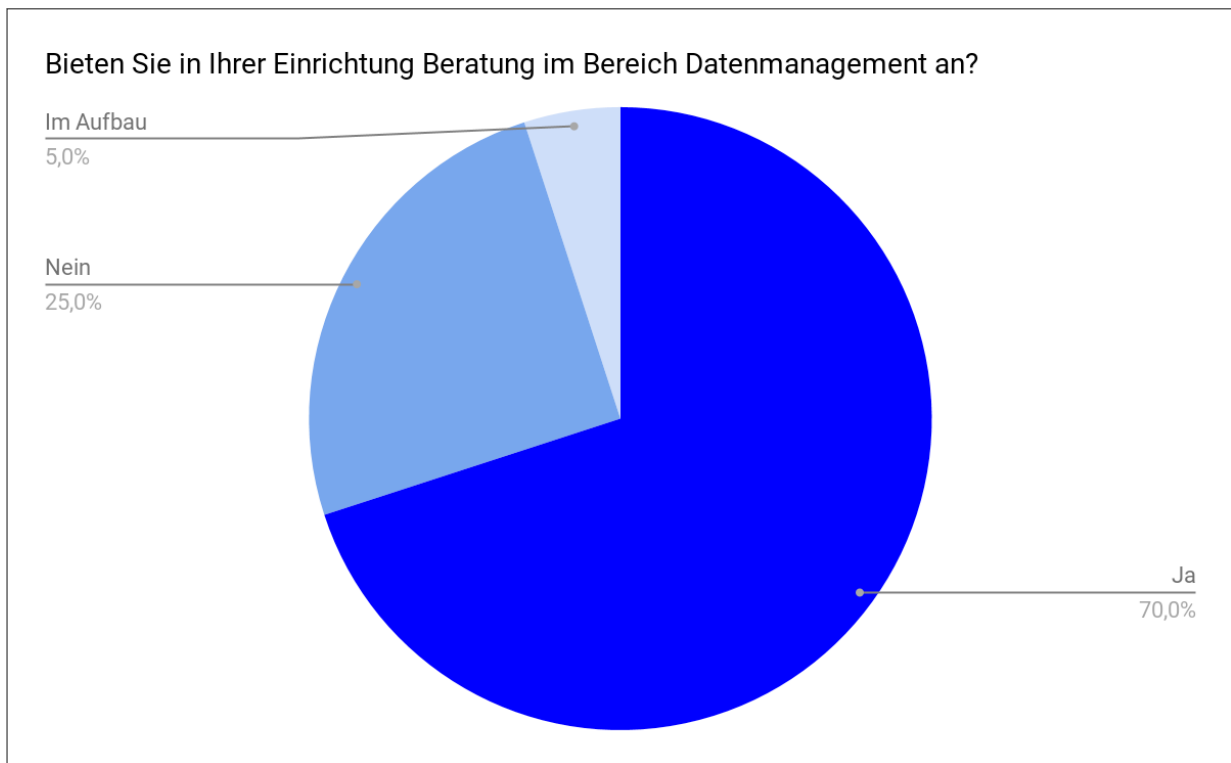
Vergleich: nicht möglich, da diese Frage im ersten Projektteil nicht gestellt wurde.



1.11. Bieten Sie in Ihrer Einrichtung Beratung im Bereich Datenmanagement an?

An 70% der Einrichtungen kann eine einschlägige Beratung, häufig auf individueller Basis, in Anspruch genommen werden. Die Formen des Angebots sind vielfältig: Schaffung eines eigenen Zentrums für Forschungsdatenmanagement; Beratung durch die forschungsunterstützenden Services; spezielle Webseiten mit umfangreichem Informationsangebot; die Bibliothek als Beratungszentrum; regelmäßige Workshops.

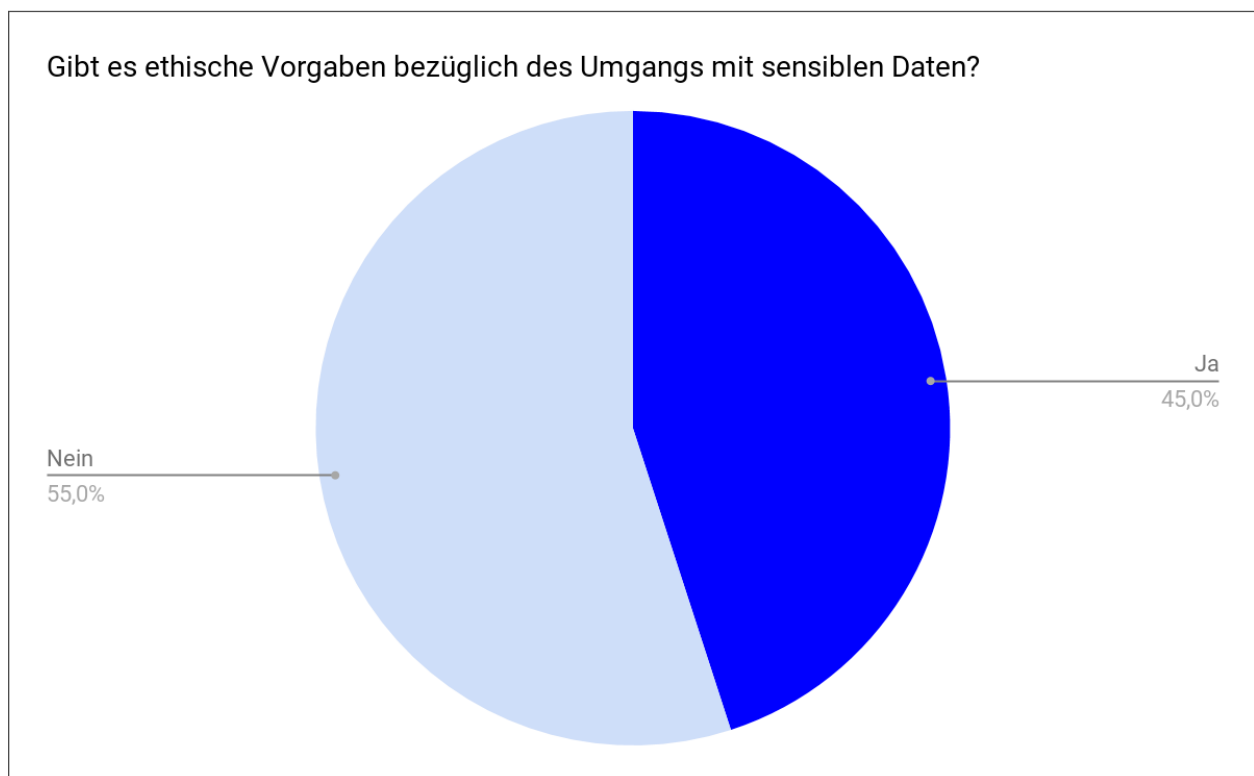
Vergleich: nicht möglich, da diese Frage im ersten Projektteil nicht gestellt wurde.



1.12. Gibt es ethische Vorgaben bezüglich des Umgangs mit sensiblen Daten? Wenn ja, welche? (Beispiele: die Möglichkeit zum Verbergen der Metadaten, Embargo, Safe Room)

Etwa die Hälfte der Einrichtung bietet in ihren Repositorien die Möglichkeit, Sperrfristen für die Freigabe von Publikationen (zumeist Hochschulschriften) gemäß § 43d des österreichischen Mediengesetzes in Anspruch zu nehmen⁴. Weitere Maßnahmen für den Schutz sensibler Daten gibt es mit einer Ausnahme (Safe Room) nicht – mit der Begründung, dass sensible Daten gar nicht erst in die Repositorien aufgenommen würden.

Vergleich: nicht möglich, da diese Frage im ersten Projektteil nicht gestellt wurde.



⁴ Da der Upload von Dissertationen etc. ins Repositorium i.d.R. nicht verpflichtend ist, ist das Schutzrecht der UrheberInnen immer gewährleistet.

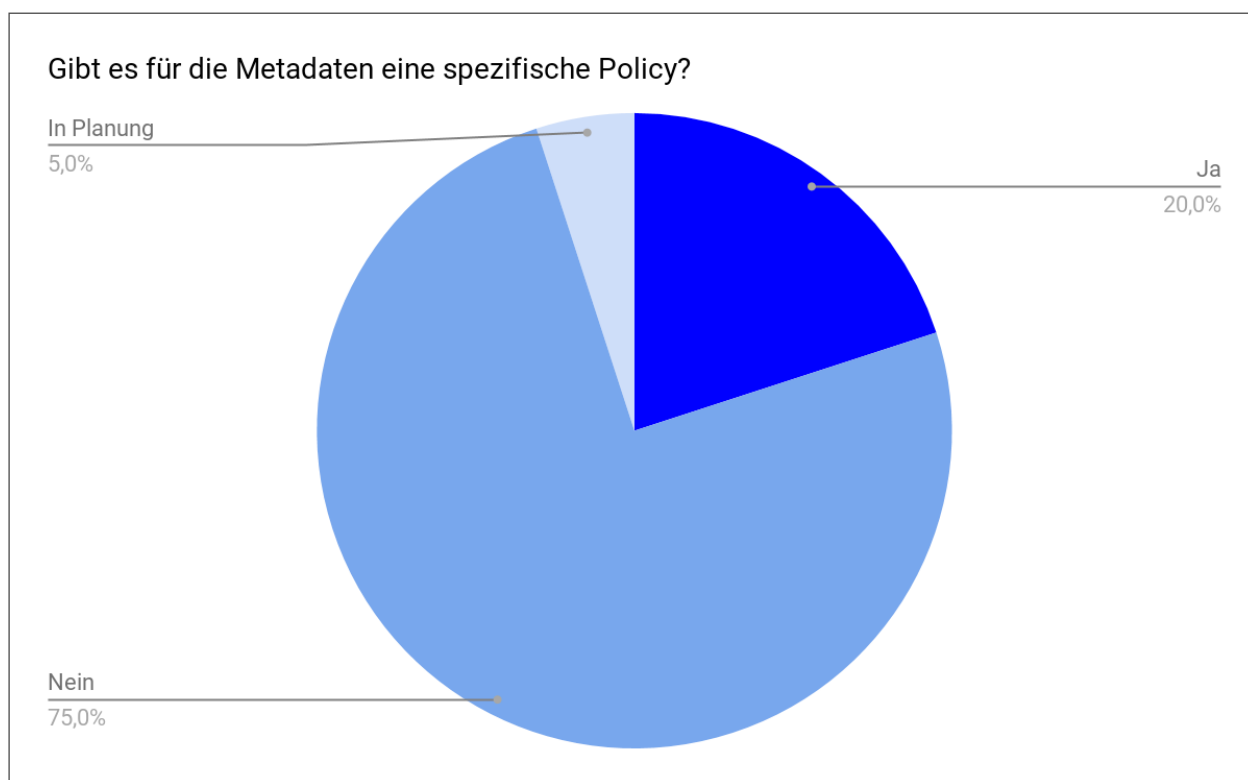
2. Fragen mit Bezug auf Metadaten

2.1. Gibt es für die Metadaten eine spezifische Policy?

Der überwiegende Anteil der Einrichtungen verfügt über keine explizit ausformulierte Metadaten-Policy. Allerdings wurde in einigen Fällen eine „implizite“ Policy angesprochen, die sich aus den technischen Gegebenheiten – der Wahl spezifischer Metadatenstandards wie Dublin Core, MODS/METS etc. – ableiten lässt. In der Praxis scheinen diese Vorgaben eine ausreichende Orientierung zu bieten, durch die sich weitere Festlegungen erübrigen. Für die Auswertung wurde das Vorhandensein einer impliziten Policy als „Ja“ gewertet, sofern damit die Wahl eines konkreten Standards gemeint war.

Für weitere Anmerkungen zum Thema „Metadaten-Policy“ vgl. den in Fußnote 1 zitierten Bericht zur Umfrage des Jahres 2015 (Abschnitt 3.2.1., S. 257f.).

Vergleich: keine nennenswerte Änderung gegenüber 2015.

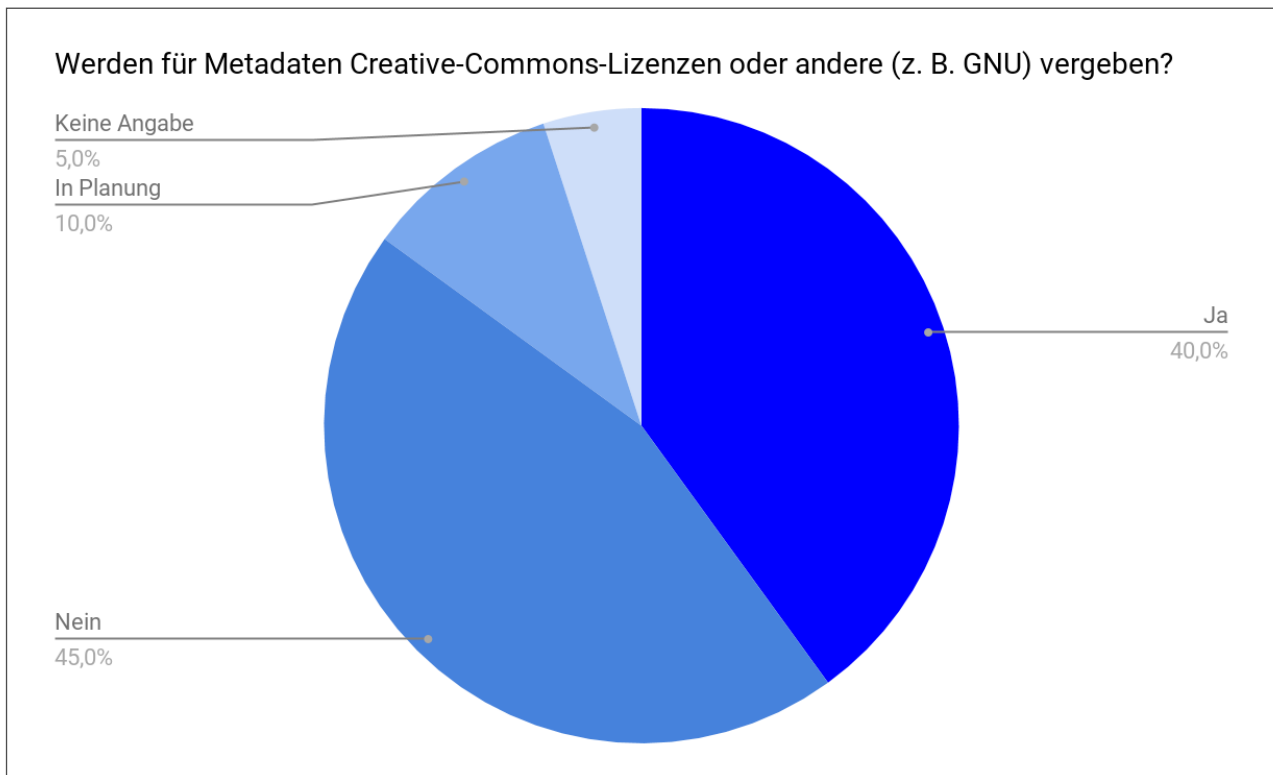


2.2. Werden für Metadaten Creative-Commons-Lizenzen oder andere (z.B. GNU) vergeben?

Rund die Hälfte der Einrichtungen ermöglicht die Vergabe von Lizenzen für Metadaten oder plant dies, zumindest in Teilbereichen (z.B. für Hochschulschriften). In einem Fall ist dies sogar verpflichtend. Als verwendetes Lizenzmodell wurde ausschließlich Creative Commons (CC) genannt.

Zu beachten ist: Nicht alle verwendeten Repositorien unterscheiden zwischen Lizenzen für Daten und Lizenzen für Metadaten – anders als beispielsweise Phaidra, wo die Lizenzen zum gegenwärtigen Stand für die Daten vergeben werden und die Metadaten frei nutzbar sind.

Vergleich: nicht möglich, da diese Frage im ersten Projektteil nicht gestellt wurde.

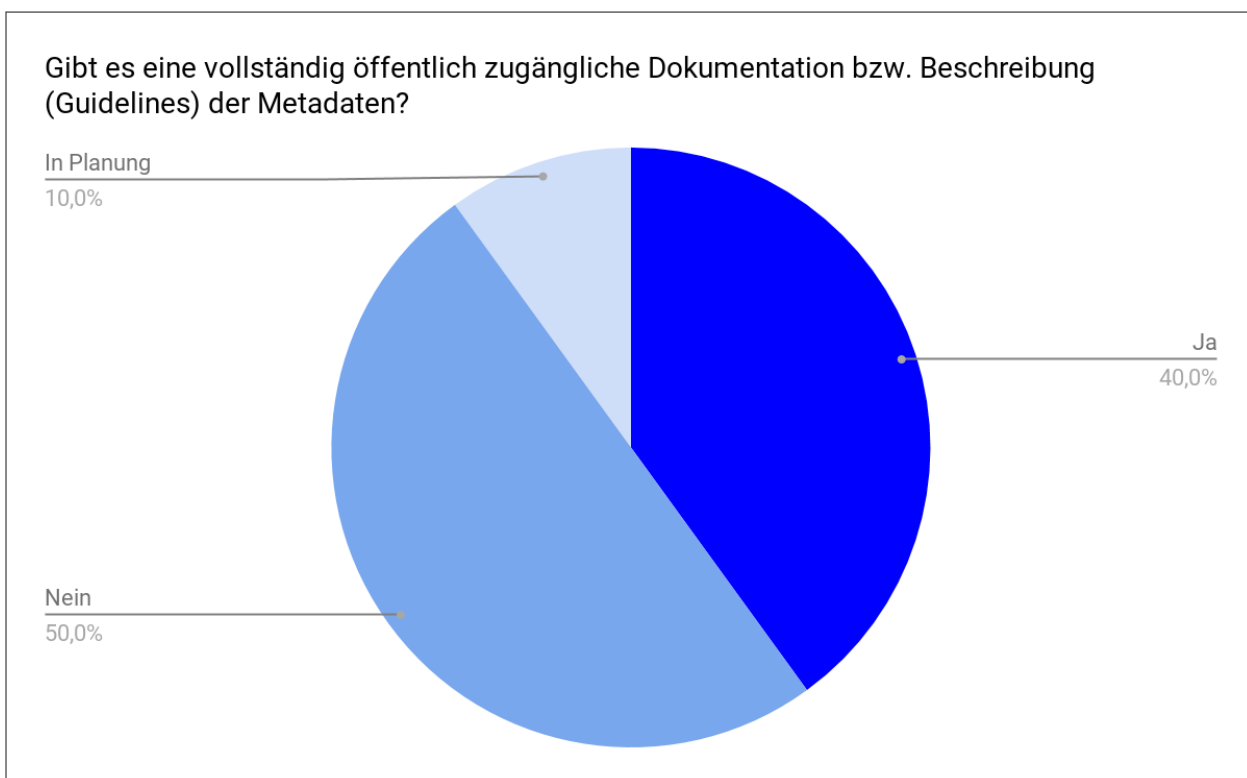


2.3. Gibt es eine vollständige öffentlich zugängliche Dokumentation bzw. Beschreibung (Guidelines) der Metadaten? Wenn nein, ist sie in Planung?

Rund die Hälfte der Einrichtungen bietet ihren NutzerInnen eine schriftliche Dokumentation zu den Metadaten an, wobei vor allem auf ins System integrierte Hilfetexte gesetzt wird. Zudem wird auf die Verwendung international etablierter Metadatenstandards mit ihrer zugehörigen Dokumentation verwiesen.

Bei der Zählung wurde die Antwort „teilweise“ als „Ja“ gewertet, noch im Aufbau befindliche Dokumentationen wurden dem Punkt „in Planung“ zugeschlagen.

Vergleich: Auch hier ergeben die ausgewerteten Daten eine Rückentwicklung: 2018 haben 50% der Institutionen keine öffentlich zugängliche Beschreibung der Metadaten, während es 2015 nur 33% waren. In einem Fall wurde die Planung aufgegeben, zweimal erfolgte ein Wechsel der Zugänglichkeit von „öffentlich“ zu „nur intern“.

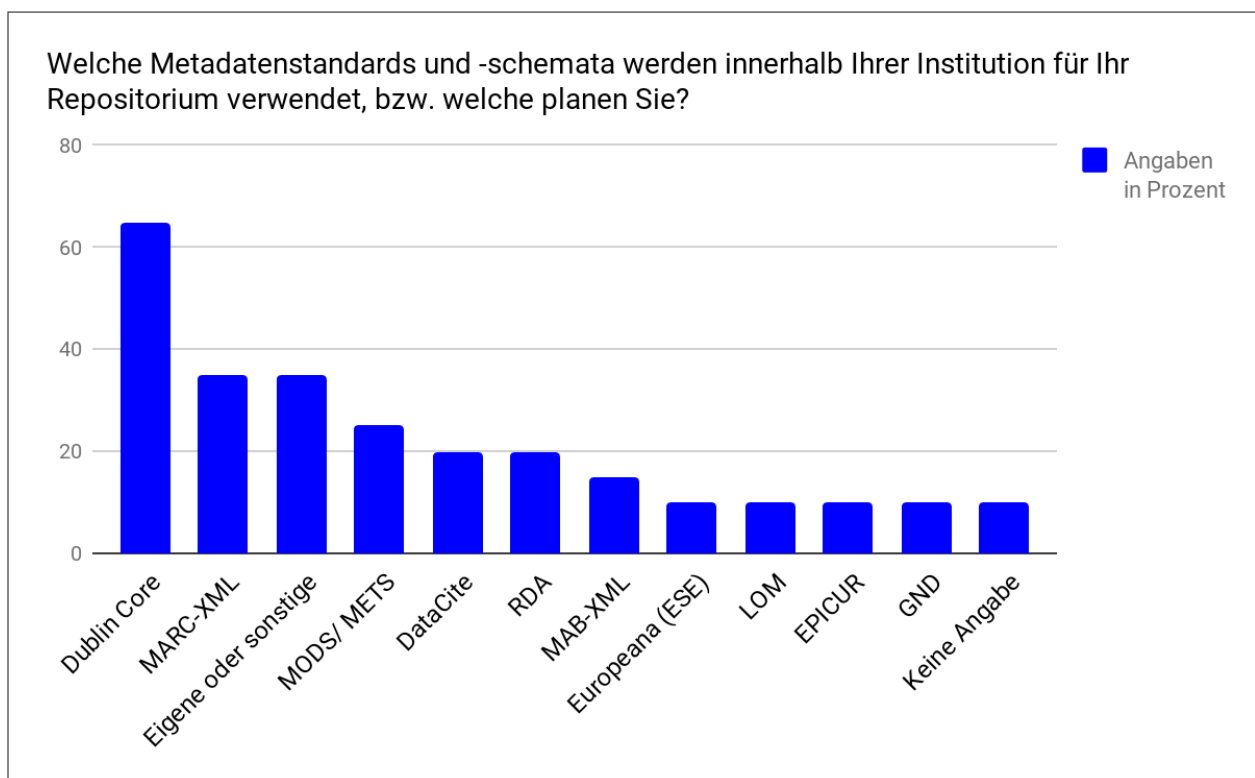


2.4. Welche Metadatenstandards und Schemata werden innerhalb Ihrer Institution für Ihr Repositorium verwendet, bzw. welche planen Sie?

Hier werden nur jene Schemata eigens berücksichtigt, die zumindest zweimal genannt wurden; die übrigen werden unter „Eigene oder sonstige“ zusammengefasst. Es gab kein vorgegebenes Antwortvokabular, was dazu führt, dass auch Metadatenstandards genannt wurden, die keine Standards für Repositorien sind. Zugelassen wurde auch die Nennung des Regelwerks RDA statt des inhaltlich besser passenden Datenformates MARC.

Die Wahl der Standards hängt natürlich stark vom jeweiligen Langzeitarchivierungssystem ab. Die wichtigsten Standards sind Dublin Core, MARC-XML und MODS/METS.

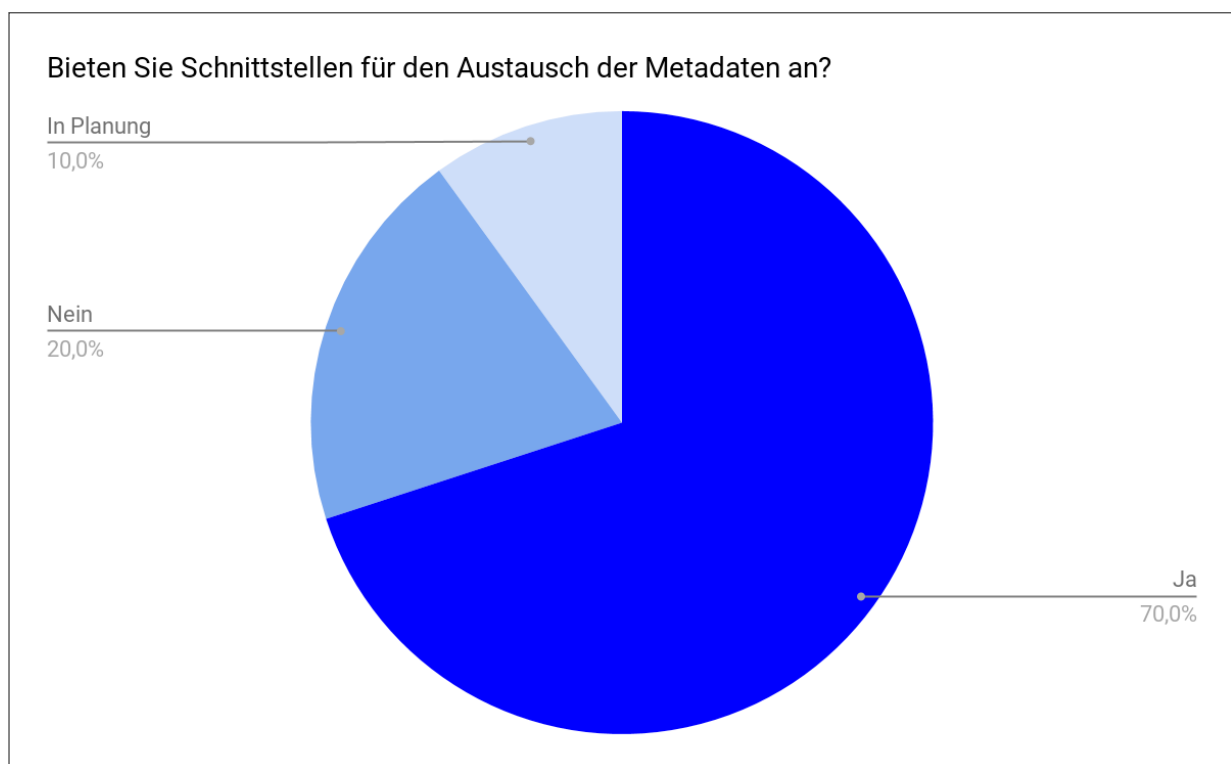
Vergleich: Dublin Core ist an die Spitze gerückt, MARC-XML (das 2015 noch gar nicht genannt wurde – eine Auswirkung des Wechsels von ALEPH zu Alma) an die zweite Stelle. Auffällig ist, dass der Datenpunkt „Andere“ – 2015 noch Spitzenreiter – 28 Prozentpunkte verloren hat. Dies ist als Bereinigung und Konzentration auf weniger, dafür gebräuchlichere Metadatenformate zu interpretieren.



2.5. Bieten Sie Schnittstellen für den Austausch der Metadaten an? Wenn ja, welche?

80% Prozent der Einrichtungen bieten eine Schnittstelle an oder haben eine solche in Planung. Der wichtigste Aspekt ist dabei die Öffnung des eigenen Repositoriums für das Metadaten-Harvesting durch andere Aggregatoren, wie z.B. die Open Archives Initiative (OAI)⁵, das Directory of Open Access Journals (DOAJ)⁶ oder das EBSCO Discovery Service (EDS). Von hoher Wichtigkeit ist natürlich auch die Schnittstelle zu Bibliothekskatalog Alma (vgl. Abschnitt 1.7.), die aber nicht als nach außen gerichtetes Angebot zu sehen ist, sondern eher als internes Werkzeug.

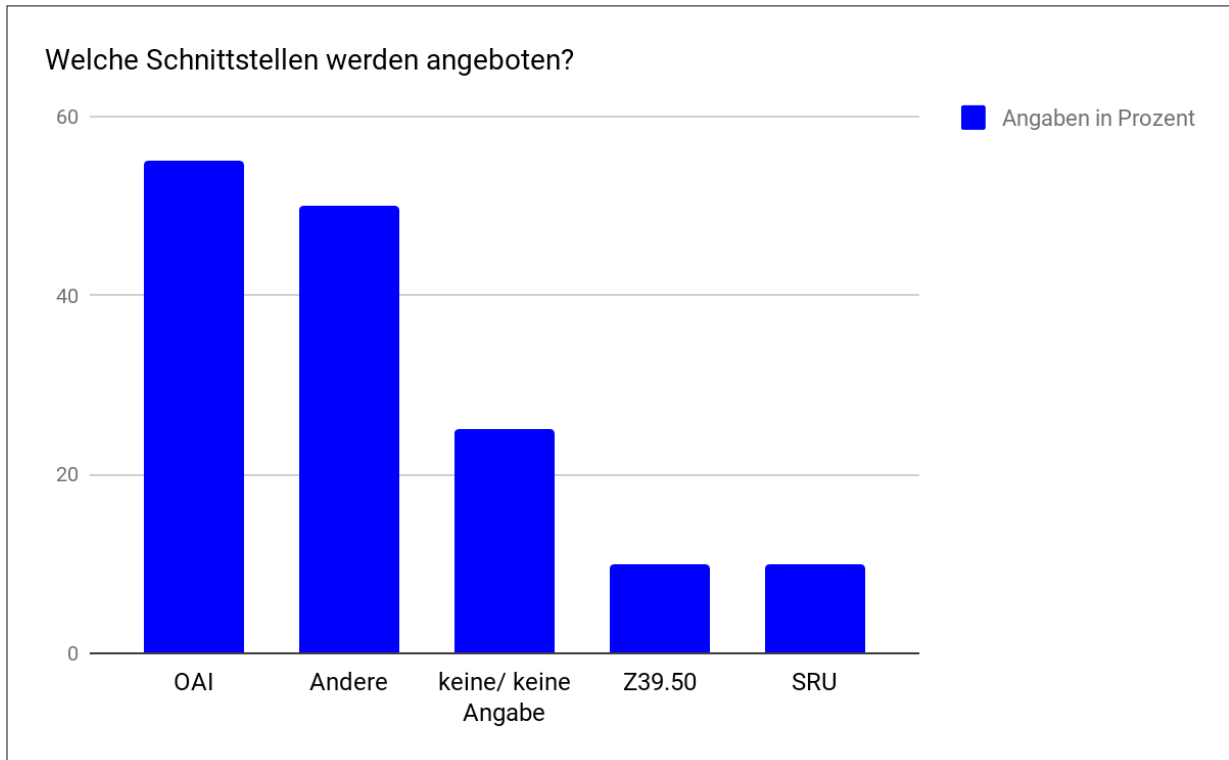
Vergleich: Ein direkter Vergleich ist nicht möglich, da diese Frage im ersten Projektteil nicht in dieser Formulierung gestellt wurde. Gefragt wurde aber (in 2.15.) ob es beabsichtigt sei, Daten und Metadaten mit anderen Einrichtungen zu teilen, was bei 71% der Befragten bereits praktiziert bzw. geplant war.



⁵ Siehe <http://www.openarchives.org/>

⁶ Siehe <https://doaj.org/>

OAI-Schnittstellen kommen an 55% der Einrichtungen zum Einsatz; kein anderes Schnittstellenmodell konnte mehr als zwei Nennungen für sich verbuchen. Genannt wurden Z39.50, SRU (Search/Retrieve via URL), SAS SWAT (SAS Scripting Wrapper for Analytics Transfer), Linked Open Data, JSON, YAML und einige andere.

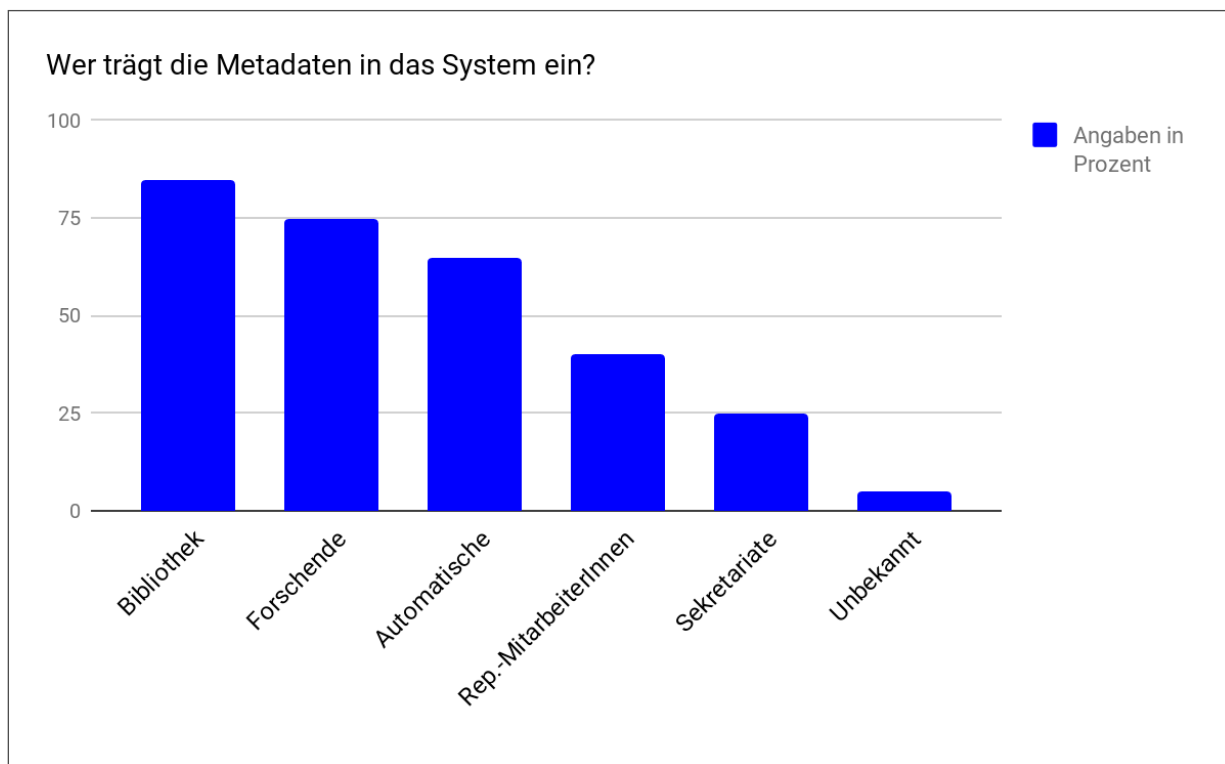


2.6. Wer trägt die Metadaten in das System ein?

Die wichtigsten MetadatenlieferantenInnen an den befragten Einrichtungen sind die Bibliotheken, knapp gefolgt von den ForscherInnen; mit etwas Abstand dahinter folgen die automatische Übernahme aus anderen Systemen und spezielle RepositoriumsmitarbeiterInnen.

Vergleich: Auch diesmal gaben 75% der Einrichtungen an, dass die Forschenden ihre Daten selbst ins Repositorium eingeben, womit dieser Anteil gleich geblieben ist. Hingegen ist bei der Antwort „BibliotheksmitarbeiterInnen“ ein Anstieg von 58% auf 85% zu verzeichnen, und bei „RepositoriumsmitarbeiterInnen“ stieg der Prozentsatz von 13% auf 40% an. Auch die automatische Übernahme ist von 29% auf 65% angewachsen.

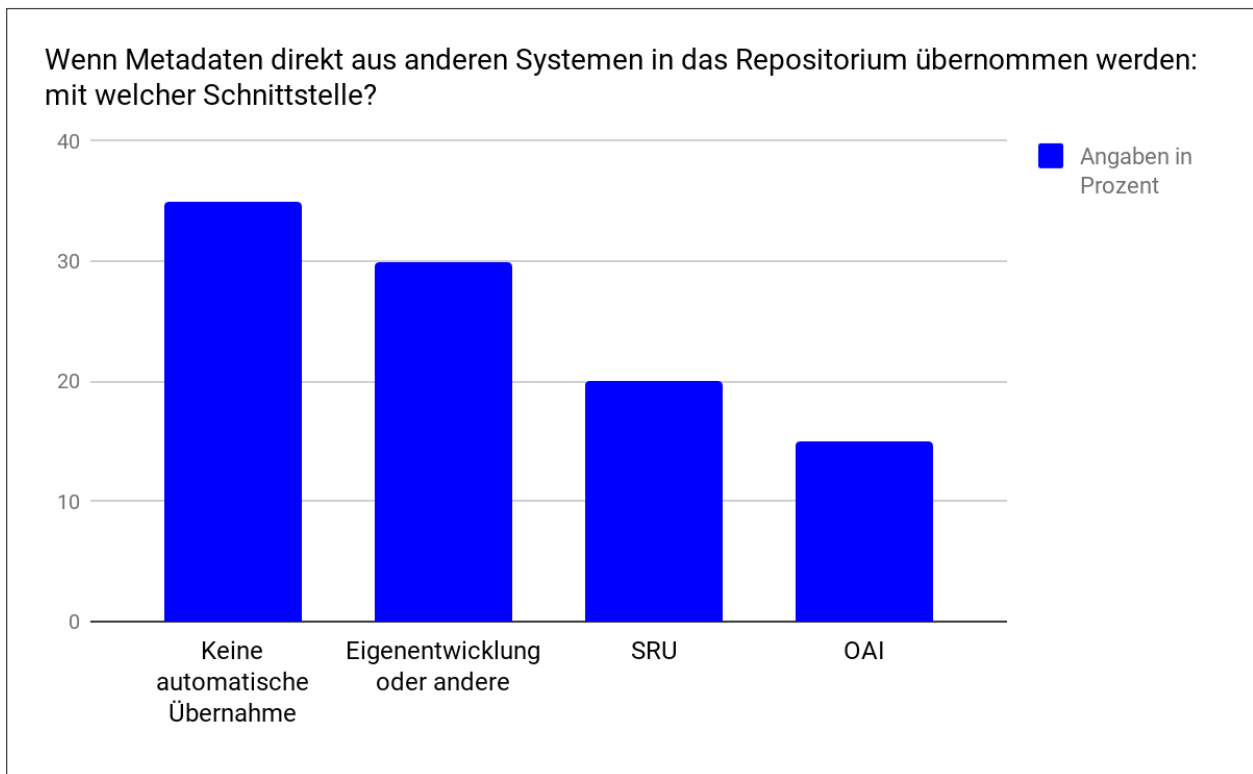
Aus diesem Vergleich lässt sich ein Trend zur Professionalisierung durch Arbeitsteilung ableiten: Während die Selbsteintragung unverändert wichtig bleibt, werden viel mehr Eintragungen als früher an andere Stellen innerhalb der jeweiligen Einrichtung (nämlich an die Bibliothek und bzw. oder an spezielles Repositoriumspersonal) ausgelagert. Auch der Automatisierungsgrad hat stark zugenommen.



2.7. Wenn Metadaten direkt aus anderen Systemen in das Repository übernommen werden: mit welcher Schnittstelle?

An 65% der Einrichtungen findet eine automatisierte Übernahme von Metadaten ins Repository statt (vgl. Abschnitt 2.6.) Viermal genannt wurde die SRU-Schnittstelle (Search / Retrieve via URL), an drei Einrichtungen setzt man auf OAI-Schnittstellen. Der Rest hat andere, z.T. auch selbst entwickelte Schnittstellen oder konnte das verwendete Schnittstellenformat nicht nennen.

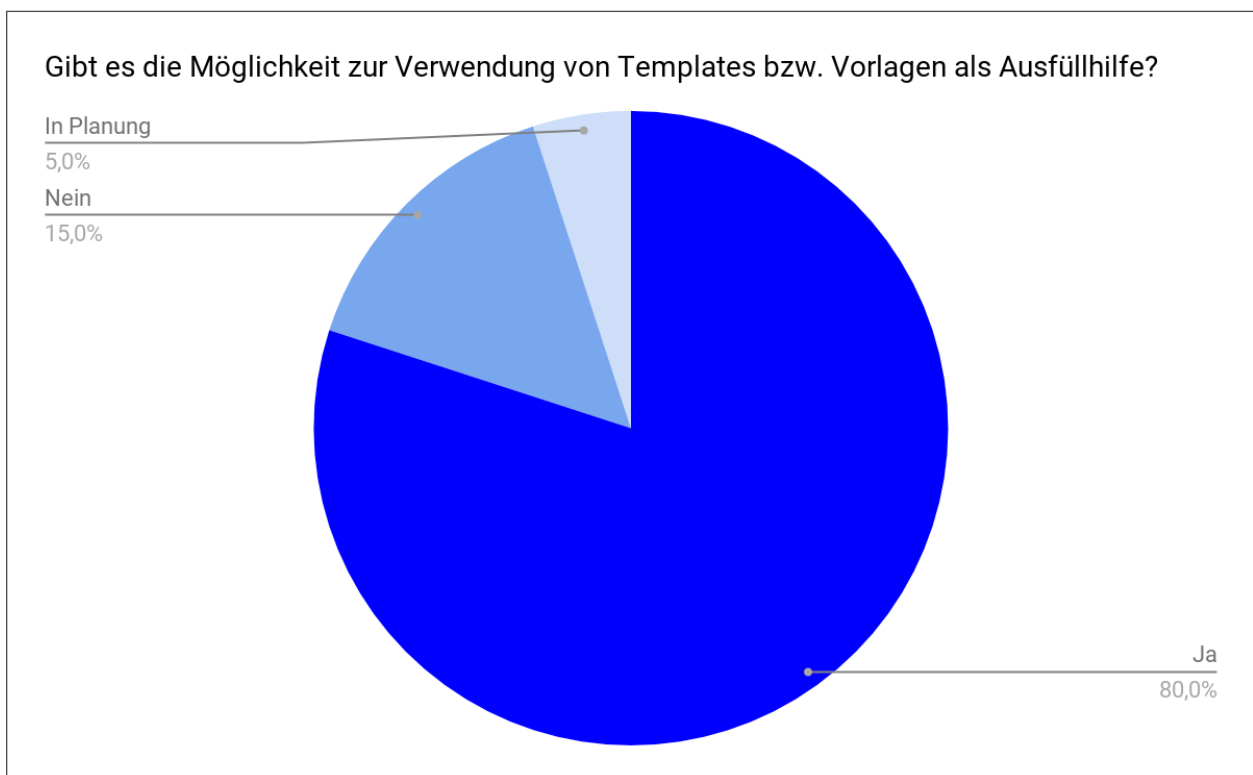
Vergleich: quantitativ nicht möglich, da im ersten Projektteil die verwendeten Schnittstellen nur in Form von Kommentaren erfasst und nicht systematisch gezählt wurden.



2.8. Gibt es die Möglichkeit zur Verwendung von Templates bzw. Vorlagen als Ausfüllhilfe?

85% der Einrichtungen bieten Ausfüllhilfen an oder haben diesbezügliche Planungen. Das Anlegen von Templates wird teilweise direkt von den Systemen unterstützt (z.B. in Phaidra); beliebt ist auch die Erstellung spezieller Eingabemasken, die einen maßgeschneiderten Zugang für die Metadatenerfassung durch nicht speziell geschulte Nutzergruppen bietet, etwa indem nur eine reduzierte, dem jeweiligen Zweck angepasste Auswahl an Metadatenfeldern angeboten wird.

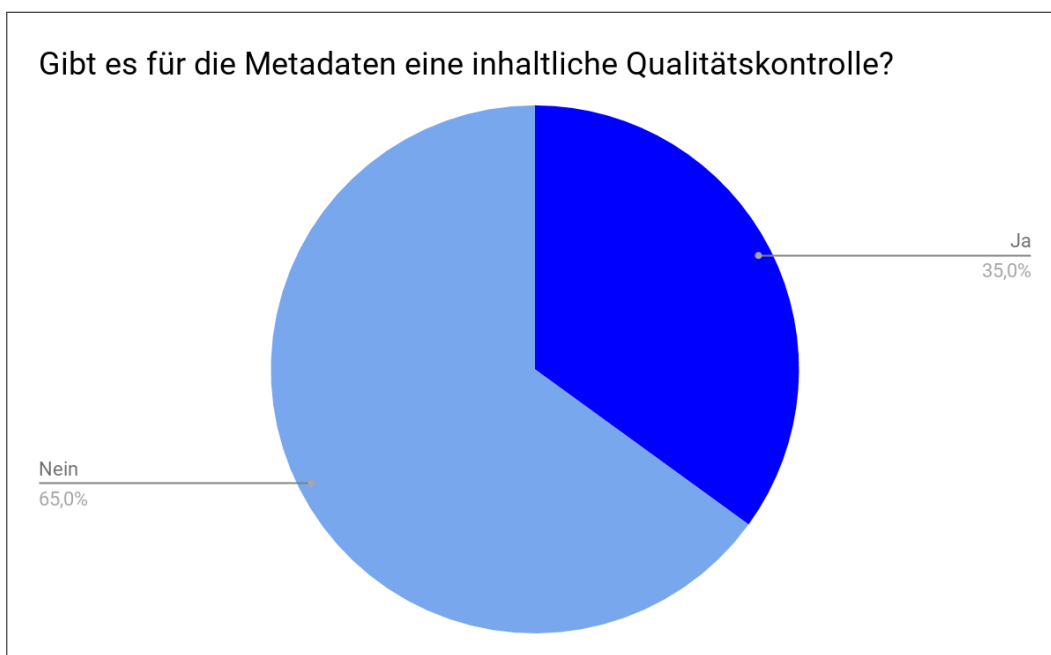
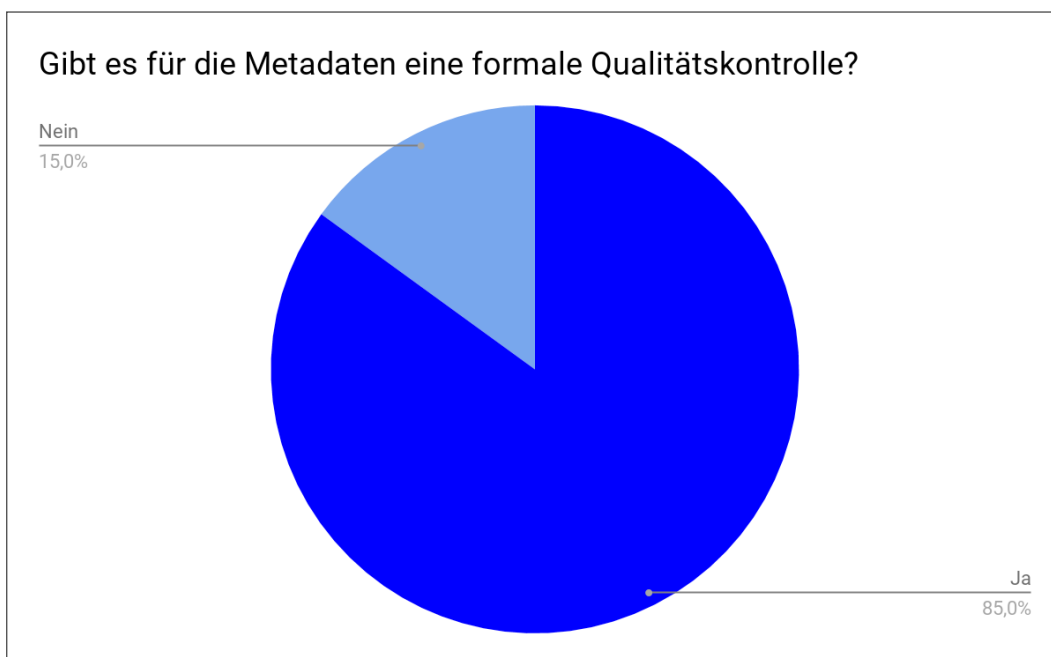
Vergleich: Gegenüber dem Jahr 2015 ist eine Steigerung von 14% – aber noch keine Sättigung – festzustellen.



2.9. Gibt es für die Metadaten eine formale und / oder inhaltliche Qualitätskontrolle?

85% der Einrichtungen nehmen eine formale, 35% auch eine inhaltliche Qualitätskontrolle der Metadaten vor. Die Methoden zur formalen Kontrolle umfassen vorwiegend Pflichtfelder in der Eingabemaske und vorgegebene Auswahllisten anstatt freier Textfelder. Die inhaltliche Kontrolle wird überwiegend von der Bibliothek der jeweiligen Einrichtung vorgenommen.

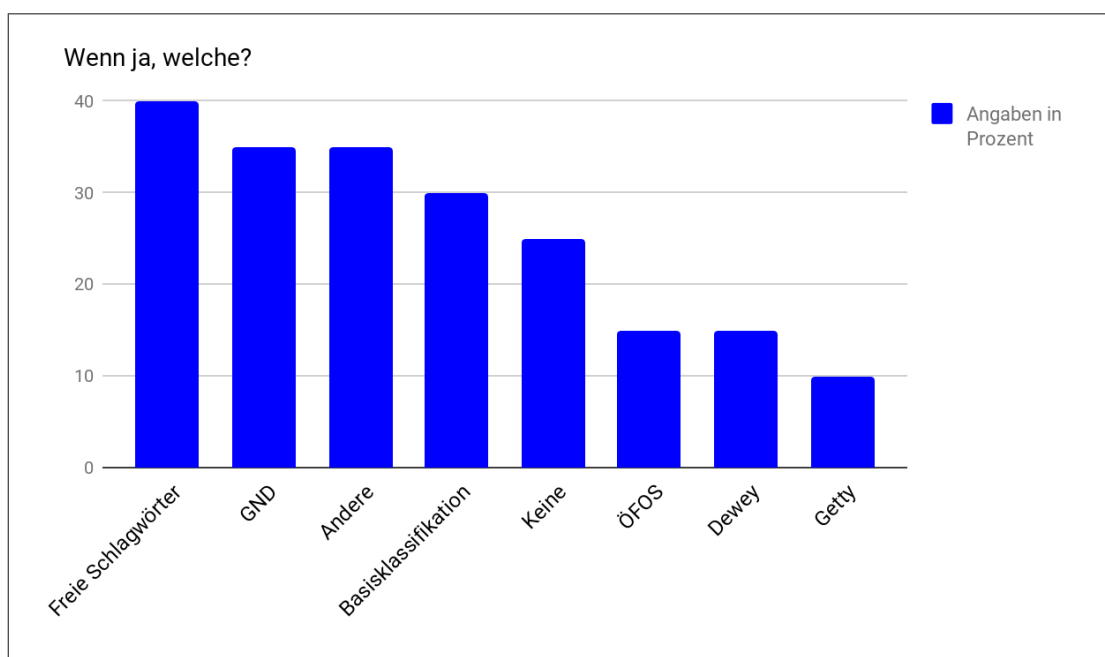
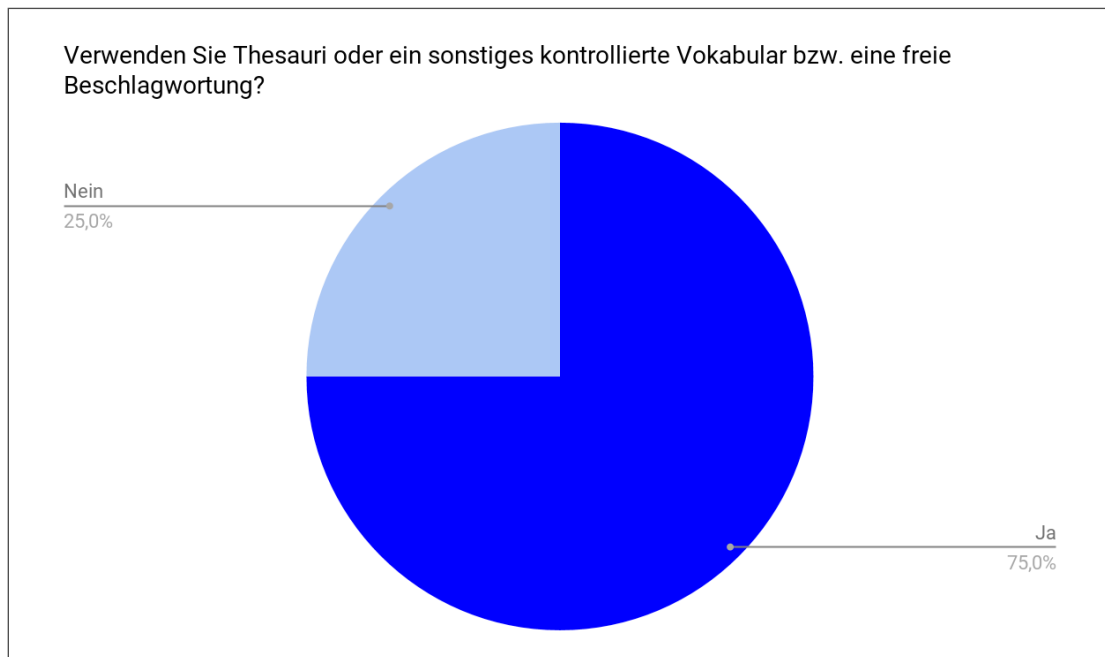
Vergleich: Bezüglich der formalen Qualitätskontrolle ist gegenüber Projektteil 1 keine nennenswerte Änderung zu vermelden. Hingegen ist der Prozentsatz der Institutionen, die auch eine inhaltliche Kontrolle durchführen, von 50% auf 35% zurückgegangen. Tatsächlich haben drei Einrichtungen, die die Frage 2015 noch mit „Ja“ beantwortet haben, diesmal auf „Nein“ umgeschwenkt, wobei die Kontrolle damals als „rudimentär“ bzw. „nur fallweise durchgeführt“ beschrieben wurde.



2.10. Verwenden Sie Thesauri oder ein sonstiges kontrolliertes Vokabular bzw. eine freie Beschlagwortung? Wenn ja, welche?

Drei Viertel der Einrichtungen bieten in ihren Repositorien die Möglichkeit zur Sacherschließung, wobei an der Spitze die freie Beschlagwortung steht, gefolgt von der Gemeinsamen Normdatei, der Basisklassifikation und – etwas abgeschlagen – ÖFOS, Dewey und die Getty-Thesauri. Jeweils nur eine Nennung bekamen IconClass, die Regensburger Verbundklassifikation (RVK), das Physics and Astronomy Classification Scheme (PACS) und einige andere.

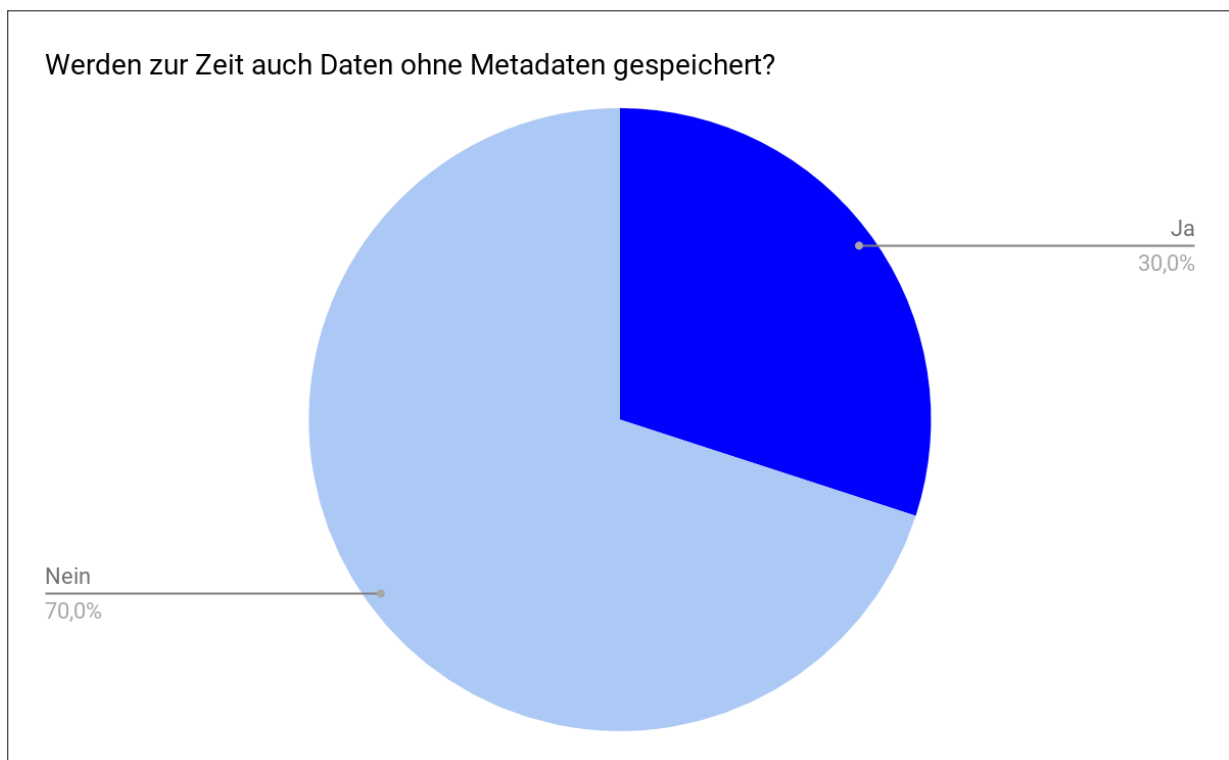
Vergleich: Bezüglich der Gesamtzahl gibt es kaum eine Änderung. Bezüglich der verwendeten Klassifikationen fällt auf, dass traditionelle Hausklassifikationen verschwunden sind; auch Dewey, Getty, ÖFOS und die Basisklassifikation haben verloren, während die GND leicht zulegen konnte.



2.11. Werden zurzeit auch Daten ohne Metadaten gespeichert?

Nur an 30% Prozent der Einrichtungen werden Daten ohne Metadaten gespeichert, wobei dies ausschließlich außerhalb der offiziellen institutionellen Repositorien so gehandhabt wird – schon allein aus dem Grund, weil die meisten Systeme dies gar nicht zulassen und ein Mindestmaß an Metadaten verlangen.

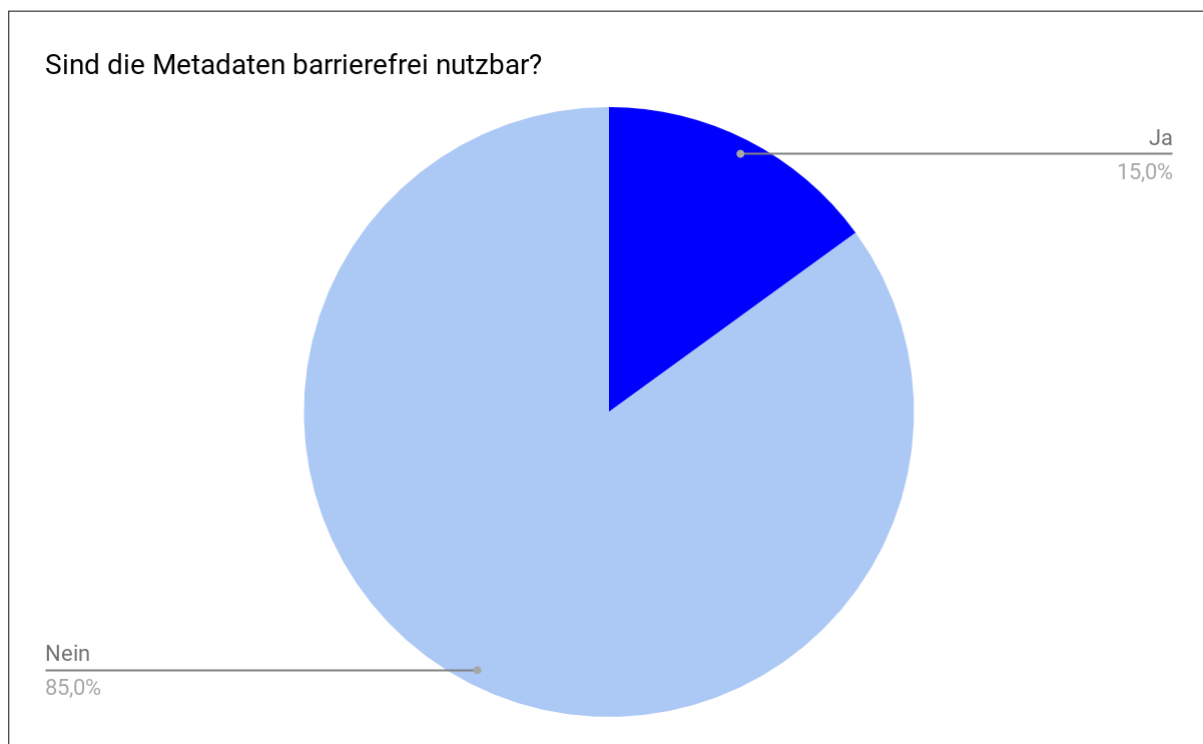
Vergleich: 2015 betrug der entsprechende Prozentsatz noch 38%, was als Hinweis darauf gesehen werden kann, dass eine Bewegung weg von „wilden“ Datensammlungen hin zu den offiziellen Repositorien stattfindet.



2.12. Sind die Metadaten barrierefrei nutzbar? Wenn ja, wie wurde das erreicht? Wenn ja, könnten Sie uns bitte ein Beispiel schicken?

Erst 15% der Einrichtungen bieten ihre Metadaten barrierefrei an. In einem Fall wird dies durch das verwendete Repositorium direkt unterstützt, in einem weiteren Fall wurde es durch speziell angepassten HTML-Quellcode bei der Anzeige erreicht und einmal ist es per Zugriff über die OAI-Schnittstelle möglich.

Vergleich: Zu beachten ist, dass 2015 nach einem barrierefreien Zugang zu den Daten gefragt wurde, was diesmal – um enger beim eigentlichen Thema der Befragung zu bleiben – in eine Frage nach den Metadaten umformuliert wurde; insofern sind die Prozentsätze nicht direkt vergleichbar. Jedenfalls haben 21% damals angegeben, Barrierefreiheit bereits zu praktizieren und 46% hatten dies geplant. Was sich jedenfalls aus den Antworten schließen lässt, ist, dass in den vergangenen drei Jahren kein großer Fortschritt erreicht wurde, denn da man für Barrierefreiheit im Gesamten genommen sowohl die Daten als auch die Metadaten diesem Prinzip unterwerfen muss, ist das Ziel erst zu einem geringen Teil erreicht.



Fazit

Insgesamt hat diese zweite Umfrage gezeigt, dass sich innerhalb der doch recht kurzen Zeit nicht nur die Repositoriumslandschaft verbessert hat, sondern auch der Umgang mit Daten und Metadaten professionalisiert werden konnte. Die Archivierung von Forschungsdaten und auch von Publikationen wurde wichtiger. Auch der mögliche Re-Use der Daten wurde stärker in den Fokus gerückt. Interessant ist, dass es in einigen Bereichen eine Rückentwicklung zu geben scheint, obwohl man bei der Interpretation der Ergebnisse von dieser doch sehr klein gehaltenen Umfrage vorsichtig sein muss. Guidelines, die früher öffentlich zugänglich waren, scheinen jetzt tendenziell eher intern genutzt werden. Ein Trend scheint die fortschreitende Automatisierung von Abläufen zu sein, auch werden mehr Templates und Ausfüllhilfen angeboten. Die inhaltliche Kontrolle der archivierten Daten hat dagegen abgenommen. Auch werden weniger Metadaten ohne Daten als 2015 archiviert, auch hier könnte man von einer Professionalisierung sprechen. Leider gilt das nicht für die Barrierefreiheit, hier ist kein nennenswerter Fortschritt zu erkennen.

Wie die Zukunft von Repositorien, vom Umgang mit Daten und Metadaten aussehen wird, lässt sich heute schwer sagen. Wichtig scheint jedoch zu sein, so eng wie möglich mit den Forschenden zusammenzuarbeiten um genau dort Unterstützung geben zu können, wo sie gebraucht wird, gleichzeitig aber auch die technischen Trends im Auge zu behalten und auf Offenheit und Flexibilität zu setzen. Wo eine Öffnung der Daten oder auch Metadaten aus ethischen oder rechtlichen Gründen nicht möglich ist, wird es neue Lösungen geben müssen. Das Thema Metadaten wird jedoch sicher nicht an Wichtigkeit verlieren, denn nur so ist die Auffindbarkeit der Daten, deren Menge weiterhin rasant zunehmen wird, auch in Zukunft gegeben.