

*„What makes it possible for the web to be composed of small pieces loosely joined?*

*Metadata“*

Jeffrey Pomerantz

---

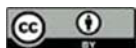
*Umgang mit Metadaten in  
Repositorien*

*Eine österreichweite Umfrage*

---

## Umgang mit Metadaten in Österreich

Work-Package-Cluster:	<b>Cluster I: Metadatenkomplex aus nicht-technischer und technischer Sicht</b>
Leitung des Clusters:	Susanne Blumesberger Universitätsbibliothek Wien <a href="mailto:susanne.blumesberger@univie.ac.at">susanne.blumesberger@univie.ac.at</a>
Datum:	12.05.2016
Version	1.1
AutorInnen:	Susanne Blumesberger  Universitätsbibliothek Wien <a href="mailto:susanne.blumesberger@univie.ac.at">susanne.blumesberger@univie.ac.at</a>  José Luis Preza  Zentraler Informatikdienst, Universität Wien <a href="mailto:jose.luis.preza@univie.ac.at">jose.luis.preza@univie.ac.at</a>  Alexander Zartl  Universitätsbibliothek Wien <a href="mailto:alexander.zartl@univie.ac.at">alexander.zartl@univie.ac.at</a>



<p>Kurzbeschreibung (Deutsch):</p>	<p>Umfrage über den Umgang mit Metadaten in digitalen Repositorien an österreichischen Forschungseinrichtungen</p>
<p>Description (English):</p>	<p>Survey about the handling of metadata in digital repositories in Austrian research institutions</p>
<p>Schlagwörter (Deutsch):</p>	<p>Metadaten – Repositorien – Österreich</p>
<p>Keywords (English):</p>	<p>Metadata – Repositories – Austria</p>





# Inhalt

<b>Liste der befragten Institutionen</b> .....	3
<b>Ergebnisse</b> .....	4
<b>1. Allgemeine Fragen</b> .....	4
1.1. Welche Arten von Daten werden an Ihrer Einrichtung gesammelt? .....	4
1.2. Gibt es bereits ein gemeinsames Archiv bzw. Repositorium für die Angehörigen Ihrer Einrichtung?.....	5
1.3. Für wie lange sollen digitale Daten aufbewahrt werden? Auf Dauer oder nur temporär? 6	
1.4. Dient die Archivierung den eigenen Forschungszielen der Einrichtung oder dem Informationsaustausch mit anderen?.....	6
1.5. Gibt es eine Schnittstelle zwischen Bibliothekskatalog und Repositorium bzw. ist eine solche in Planung?.....	7
<b>2. Fragen mit Bezug auf Metadaten</b> .....	8
2.1. Gibt es für die Metadaten eine spezifische Policy?.....	8
2.2. Gibt es eine vollständige Dokumentation bzw. Beschreibung der Metadaten?.....	8
2.3. Welche Metadatenstandards und -Schemata werden innerhalb Ihrer Institution verwendet? .....	9
2.4. Auf welcher Entscheidungsgrundlage wurden die verwendeten Schemata ausgewählt? . 9	
2.5. Wo sehen Sie die Grenzen der ausgewählten Schemata und Standards?.....	10
2.6. Ist es grundsätzlich möglich, die Schemata an andere Einrichtungen weiterzugeben? Wie wurde dies erreicht? .....	11
2.7. Ist geplant, die Metadaten-Schemata über die Einrichtung hinaus verfügbar zu machen? .....	12
2.8. Wer trägt die Metadaten ins System ein?.....	12
2.9. Gibt es zum Eintragen der Metadaten Hilfestellung (z.B. Hilfstexte, wie die Metadatenfelder auszufüllen sind)? Wo findet man diese?.....	13
2.10. Gibt es die Möglichkeit zur Verwendung von Templates als Ausfüllhilfe? .....	14
2.11. Gibt es für die Metadaten eine formale und / oder inhaltliche Qualitätskontrolle? .....	15
2.12. Verwenden Sie einen Thesaurus oder ein sonstiges kontrolliertes Vokabular? Wenn ja, welches?.....	16
2.13. Wie viele Felder für die Metadaten sind vorgesehen? .....	17
2.14. Werden zurzeit auch Daten ohne Metadaten gespeichert? .....	17
2.15. Ist es beabsichtigt, Daten und Metadaten mit anderen Einrichtungen zu teilen? .....	18
2.16. Gibt es einen barrierefreien Zugang zu den Daten? .....	18

## Vorbemerkung

Ziel der Befragung war, festzustellen, wie man möglichst rasch und effizient herausfinden könnte, wie die Partnerinstitutionen am Projekt *e-Infrastructures Austria* derzeit mit Metadaten in ihren Repositorien umgehen, bzw. wie ihre Planung diesbezüglich aussieht. Da zum Zeitpunkt der Fragebogenerstellung (April 2015) noch nicht alle Partnerinstitutionen über ein Repositoryum verfügten und somit auch keine Erfahrungen mit Metadaten in diesem Bereich hatten, musste die Befragung ziemlich niederschwellig angesetzt werden. Die Gruppe einigte sich auf einen schriftlichen Fragebogen, der zwar an die AnsprechpartnerInnen geschickt werden sollte, von ihnen jedoch nicht alleine ausgefüllt, sondern in Form eines angekündigten telefonischen Gesprächs abgehandelt werden sollte. Der Vorteil dabei ist, dass im persönlichen Gespräch nachgehakt werden kann und sollte etwas nicht oder falsch verstanden worden sein, die Möglichkeit zur Nachfrage gegeben ist.

In einem ersten Schritt wurden die Fragen überlegt und möglichst klar und deutlich formuliert. Danach wurden die jeweiligen AnsprechpartnerInnen an den einzelnen Institutionen gesucht. In einer vorbereiteten Mail wurde um einen Gesprächstermin ersucht, der Fragebogen wurde bereits beim ersten Kontakt mitgeschickt, um den GesprächspartnerInnen die Möglichkeit zu geben, sich vorab Antworten überlegen zu können, bzw. auch um ihnen die Gelegenheit zu bieten, weitere ExpertInnen an der eigenen Institution in die Befragung miteinzubeziehen.

Während des telefonischen Interviews wurde das Gespräch in Stichworten niedergeschrieben und gleich danach ergänzt. Die personalisierten Fragebögen wurden archiviert, aber – so ist es mit den InterviewpartnerInnen abgesprochen – nicht an die Öffentlichkeit weitergegeben, sondern nur anonym weiter verarbeitet.

Zu den Umfrageergebnissen ist zu ergänzen, dass die Ergebnisse natürlich nur eine Momentaufnahme sein können. Es zeigt sich jedoch bereits deutlich, dass der Umgang mit Metadaten in einem Repositoryum noch keineswegs selbstverständlich bzw. erprobt ist. Für viele Institutionen sind Repositorien eine neue Aufgabe. Metadaten sind zwar aus dem bibliothekarischen Alltag bekannt, die Umsetzung dieses Wissens auf die Beschreibung digitaler Objekte ist jedoch noch relativ ungewohnt und steht ziemlich am Anfang. Hier ist nicht nur technisches Know-How wichtig, sondern vor allem auch das Bewusstsein, wie wichtig Metadaten für ForscherInnen bzw. für künftige NutzerInnen der Objekte sind.

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.

## Liste der befragten Institutionen

Mit Ausnahme der Österreichischen Bibliothekenverbund und Service Ges.m.b.H., die aufgrund ihrer speziellen Aufgabenstellung mit Bezug auf den gesamten Bibliotheksbereich eine Sonderstellung einnimmt und daher nicht vergleichbar ist, konnten im Zeitraum von Juni bis September 2015 alle am Projekt *e-Infrastructures Austria* teilnehmenden Organisationen in die Befragung einbezogen werden.

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

# Ergebnisse

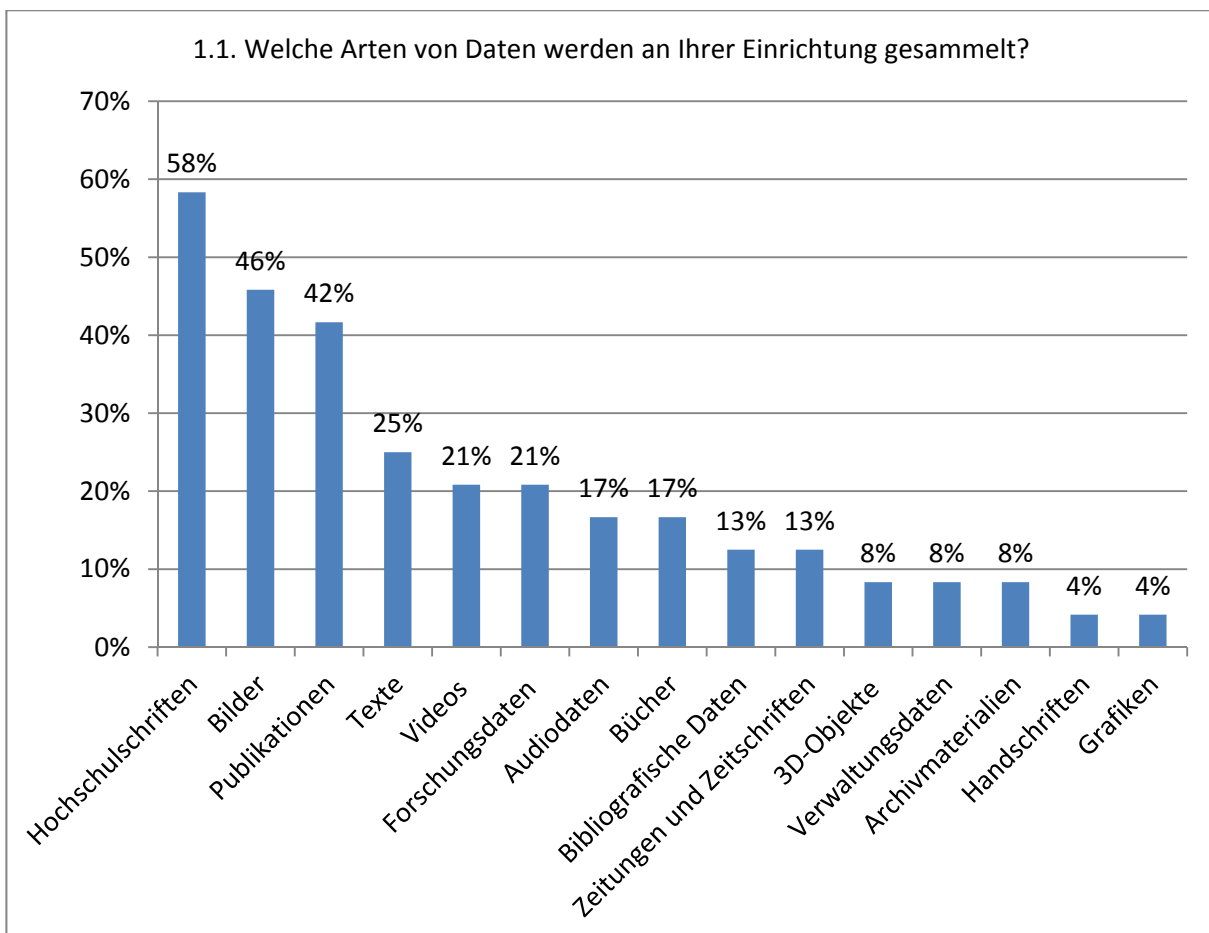
## 1. Allgemeine Fragen

### 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.

Hier wird ersichtlich, dass an erster Stelle Hochschulschriften archiviert werden (sollen), in weiterer Folge Bilder, Publikationen, Texte, Videos, und an 6. Stelle Forschungsdaten.

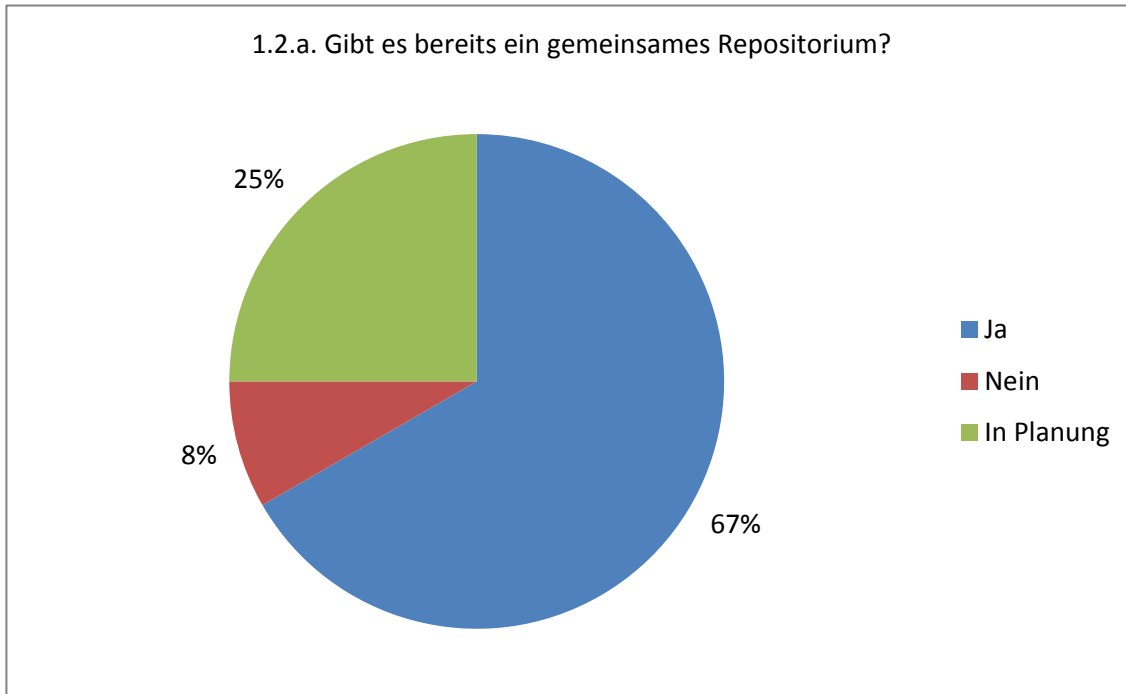
Speziell zum Begriff „Forschungsdaten“ ist anzumerken, dass diese Bezeichnung weiter ausgelegt wurde, als es die üblichen Definition – Messprotokolle, Umfragedaten, sonstige Rohdaten – nahelegt. So wurden z.B. in einem Fall explizit multimediale (Video-)Daten darunter verstanden. Insgesamt gilt damit das Prinzip, dass der Kontext bestimmt, was Forschungsdaten sind und nicht der Datentyp.



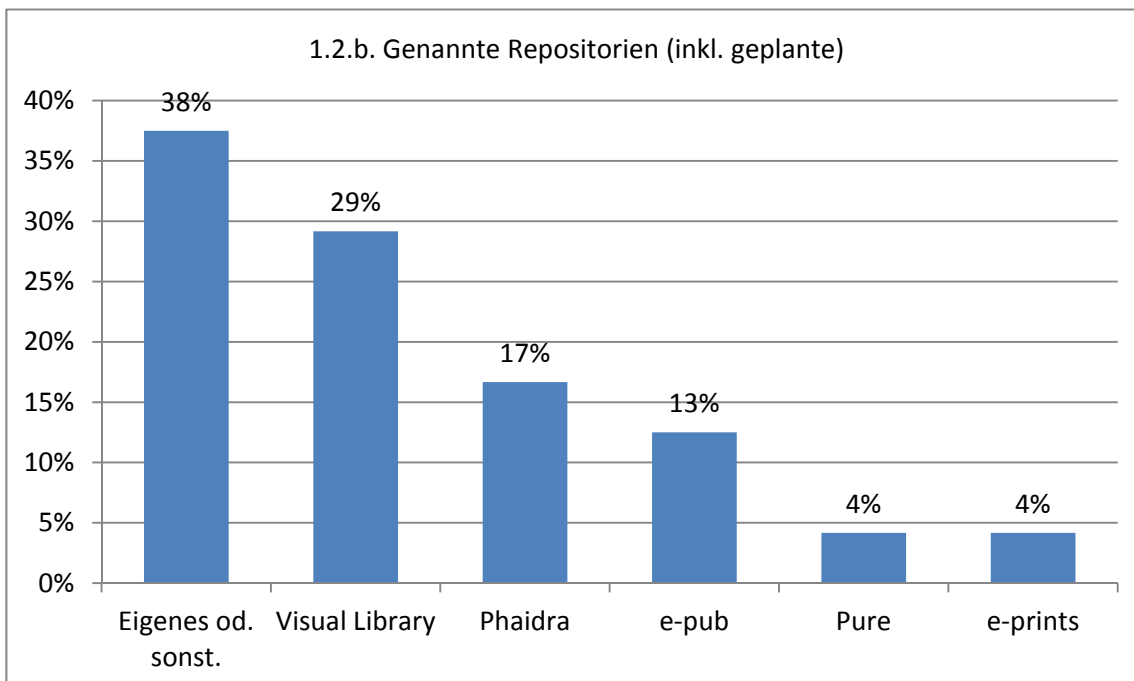


### 1.2. Gibt es bereits ein gemeinsames Archiv bzw. Repositorium für die Angehörigen Ihrer Einrichtung?

Nur zwei der TeilnehmerInnen am Projekt e-Infrastructures Austria planten zum Zeitpunkt der Befragung noch kein Repositorium; bei den in Planung befindlichen Repositorien war noch nicht überall bekannt, welches System gewählt werden würde.



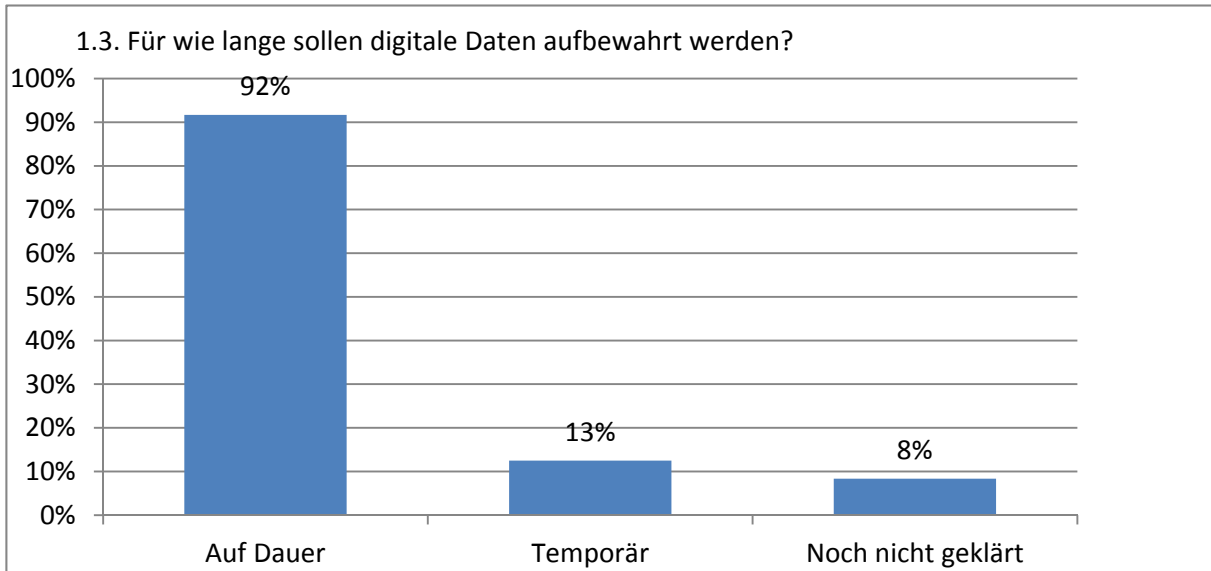
Bei den bereits vorhandenen oder fix geplanten Systemen sah die Verteilung folgendermaßen aus:



Abgesehen von den Spitzenreitern Visual Library, Phaidra und e-pub, die an mehr als zehn Prozent der Einrichtungen in Gebrauch stehen, zeigt sich eine sehr uneinheitliche Repositorienlandschaft mit einem hohen Anteil an Eigenentwicklungen.

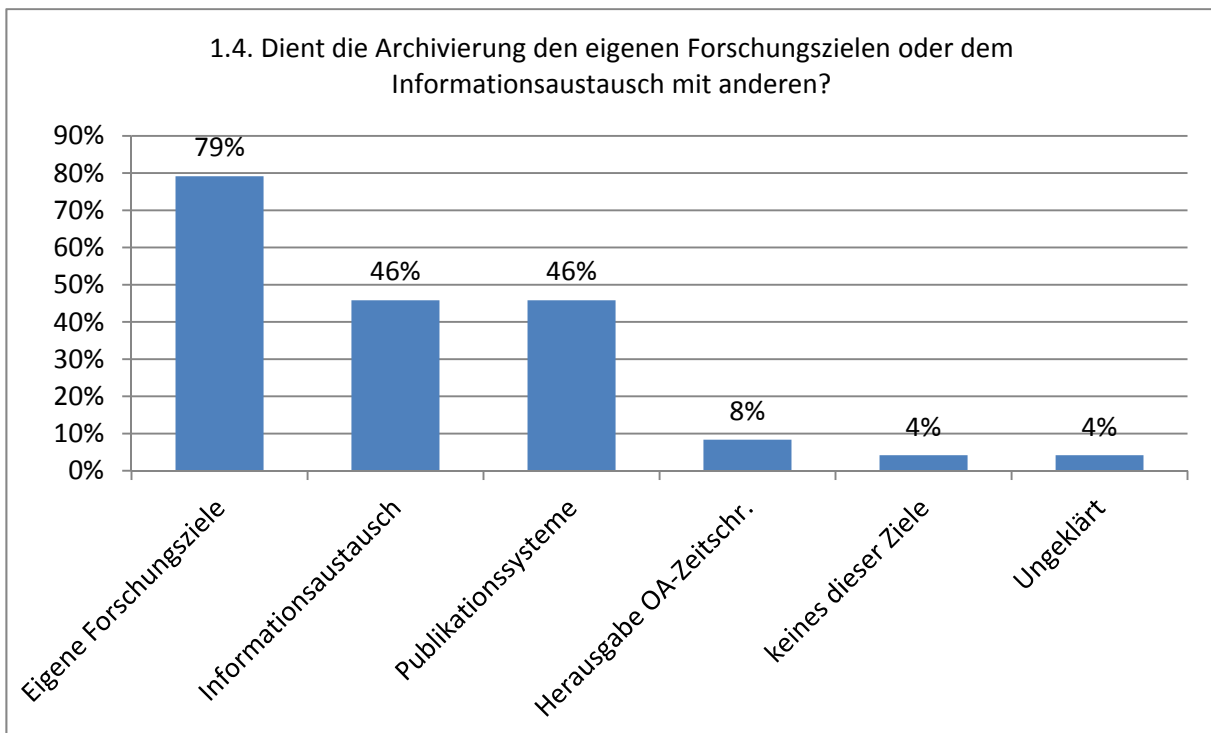
### 1.3. Für wie lange sollen digitale Daten aufbewahrt werden? Auf Dauer oder nur temporär?

Die Grafik zeigt, dass alle befragten Institutionen, die diese Frage schon geklärt haben, ihre Daten langzeitarchivieren möchten. In drei Fällen ist zusätzlich auch die Möglichkeit der temporären Speicherung wichtig. Der Wunsch nach rein temporärer Archivierung wurde nicht genannt.



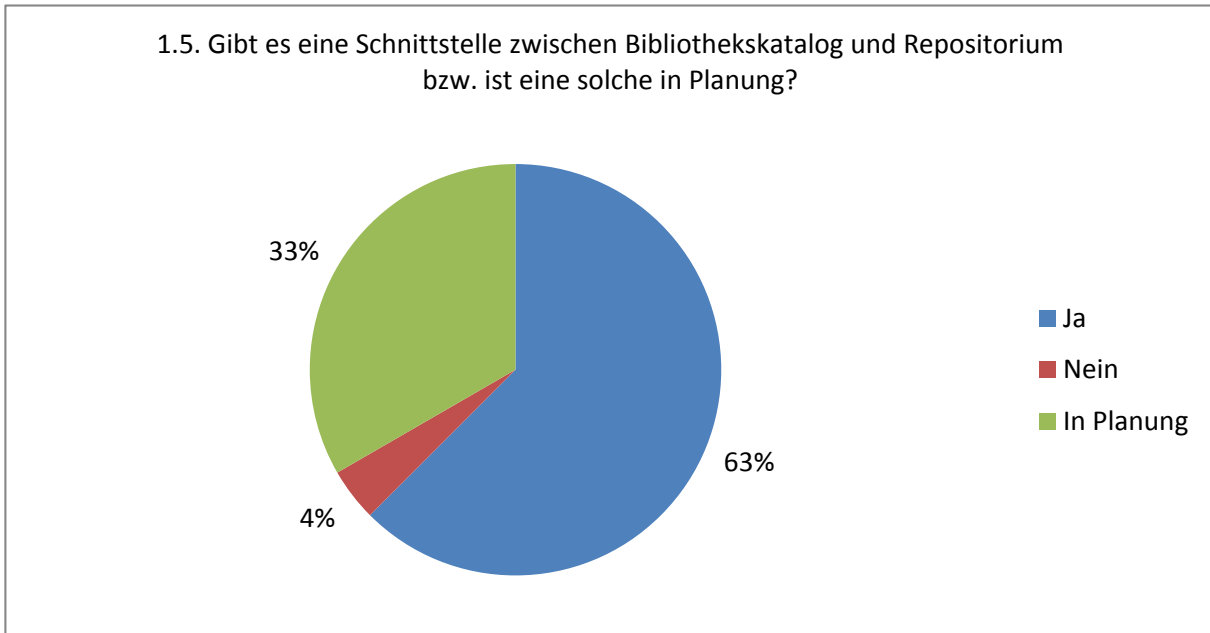
### 1.4. Dient die Archivierung den eigenen Forschungszielen der Einrichtung oder dem Informationsaustausch mit anderen?

Bei dieser Frage waren mehrere Antworten möglich. Die Grafik zeigt, dass die Archivierung hauptsächlich der eigenen Forschung, bzw. der Darstellung des eigenen Forschungsoutputs dient. An der zweiten Stelle steht der Informationsaustausch.



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

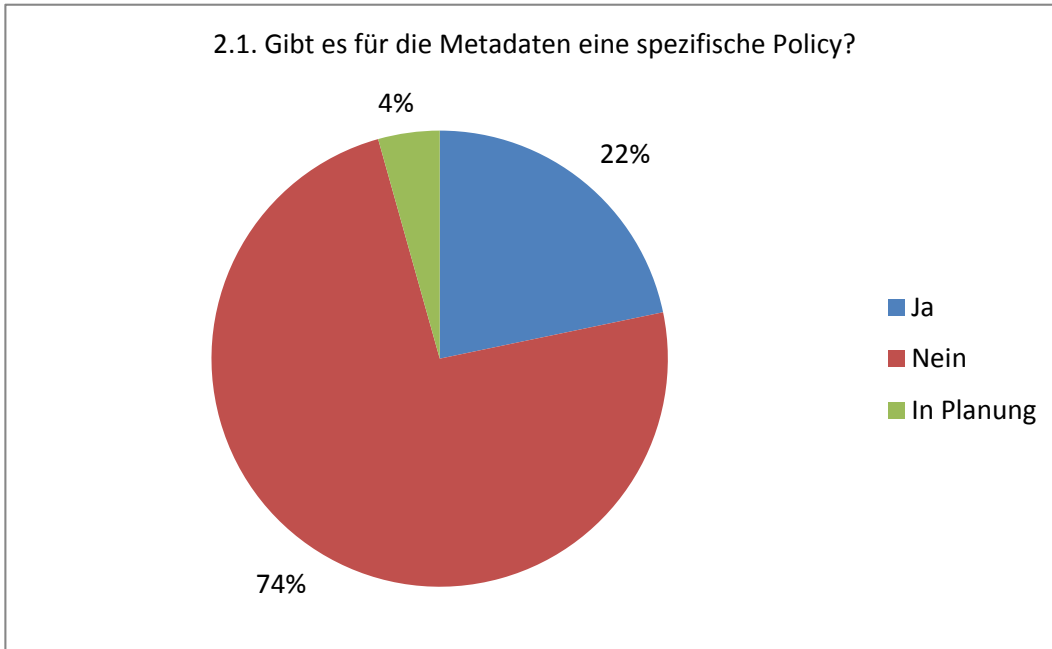
Die Schnittstelle zwischen Bibliothekskatalog und Repository ist für die meisten Partner bereits realisiert oder in Planung.



## 2. Fragen mit Bezug auf Metadaten

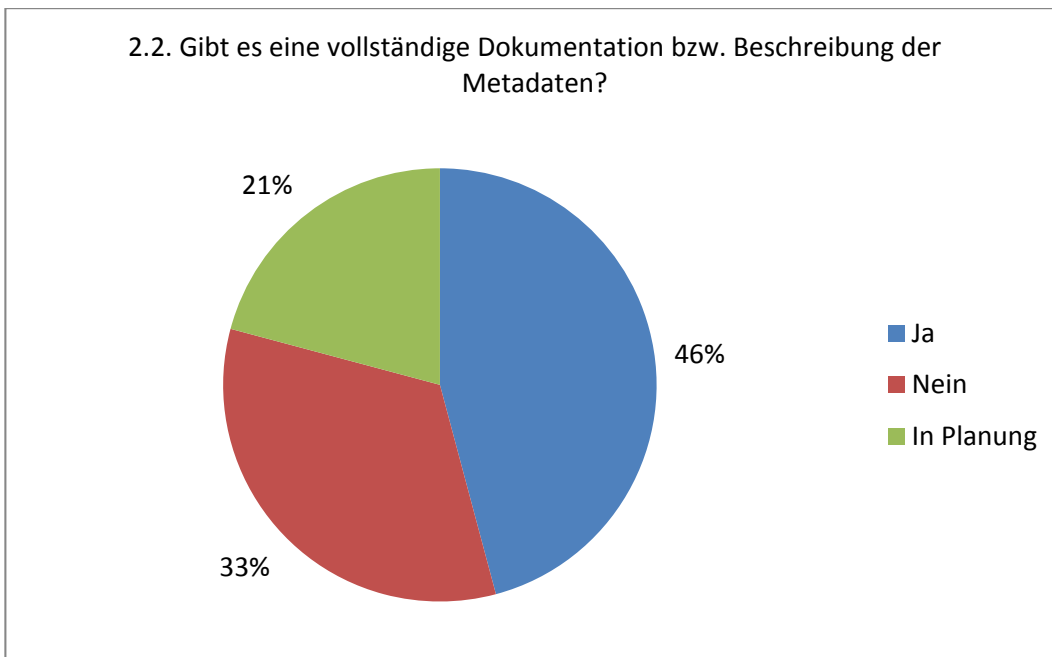
### 2.1. Gibt es für die Metadaten eine spezifische Policy?

Wie ersichtlich, haben erst sehr wenige Institutionen eine spezifisch ausformulierte Policy für Metadaten. Jene, die bereits eine Policy haben, verwenden i.d.R. internationale Standards.



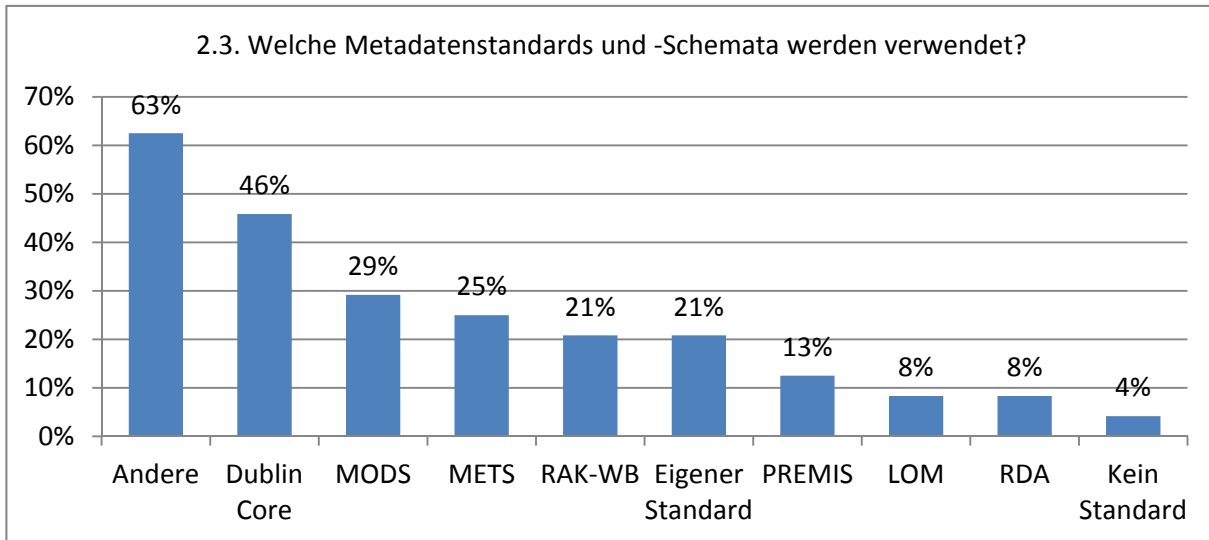
### 2.2. Gibt es eine vollständige Dokumentation bzw. Beschreibung der Metadaten?

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



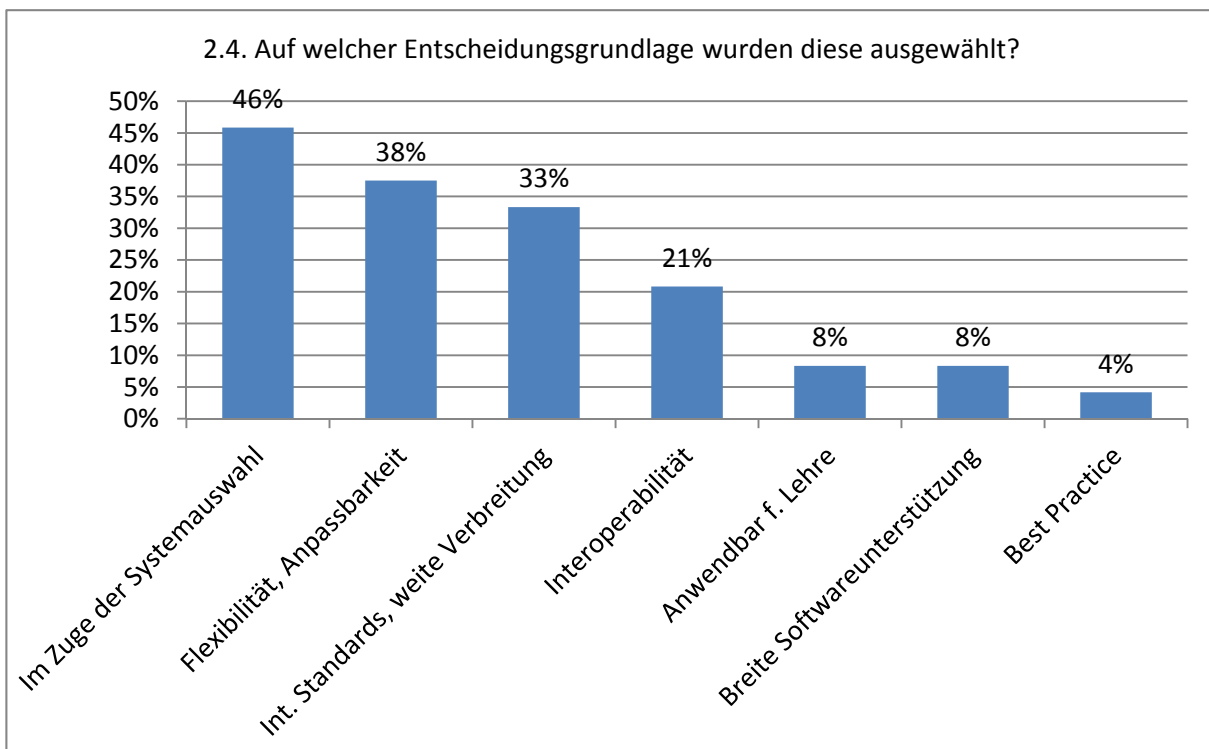
### 2.3. Welche Metadatenstandards und -Schemata werden innerhalb Ihrer Institution verwendet?

Hier werden nur die am häufigsten genannten Schemata eigens berücksichtigt, die übrigen werden unter „Andere“ 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 der Regelwerke RAK bzw. RDA statt der inhaltlich besser passenden Datenformate MAB oder MARC. - Die Wahl der Standards hängt natürlich stark vom jeweiligen Langzeitarchivierungssystem ab. Dublin Core ist Spitzenreiter.



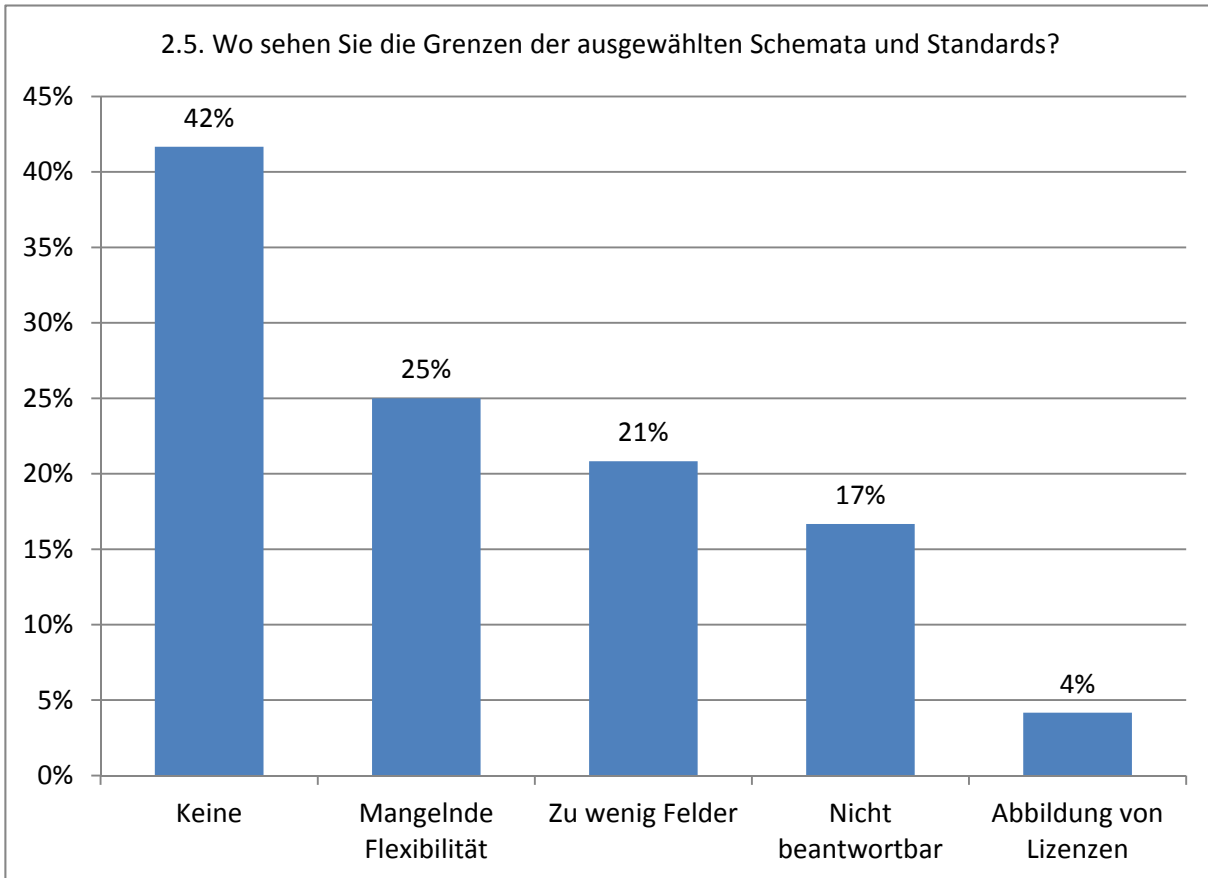
### 2.4. Auf welcher Entscheidungsgrundlage wurden die verwendeten Schemata ausgewählt?

Die Antworten waren relativ breit gefächert. An erster Stelle steht „Im Zuge der Systemauswahl“, was wohl impliziert, dass bei der Systemauswahl auch andere Kriterien als die dort verwendeten Metadatenstandards im Vordergrund standen und diese gewissermaßen „mitgekauft“ wurden. Auch auf Flexibilität und die Verwendung internationaler Standards (vgl. Frage 2.1.) wurde Wert gelegt.



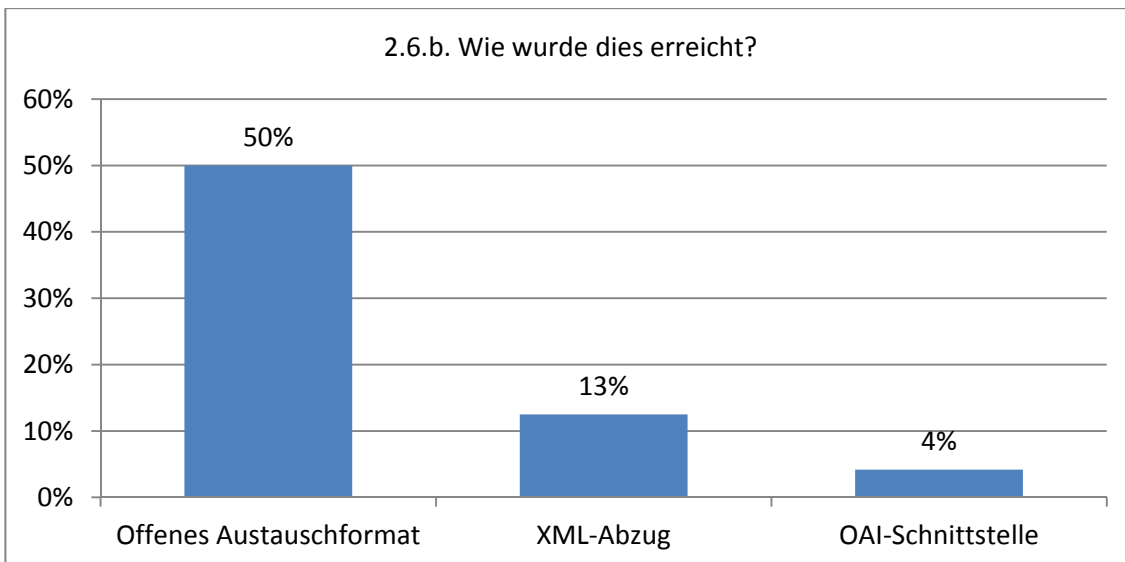
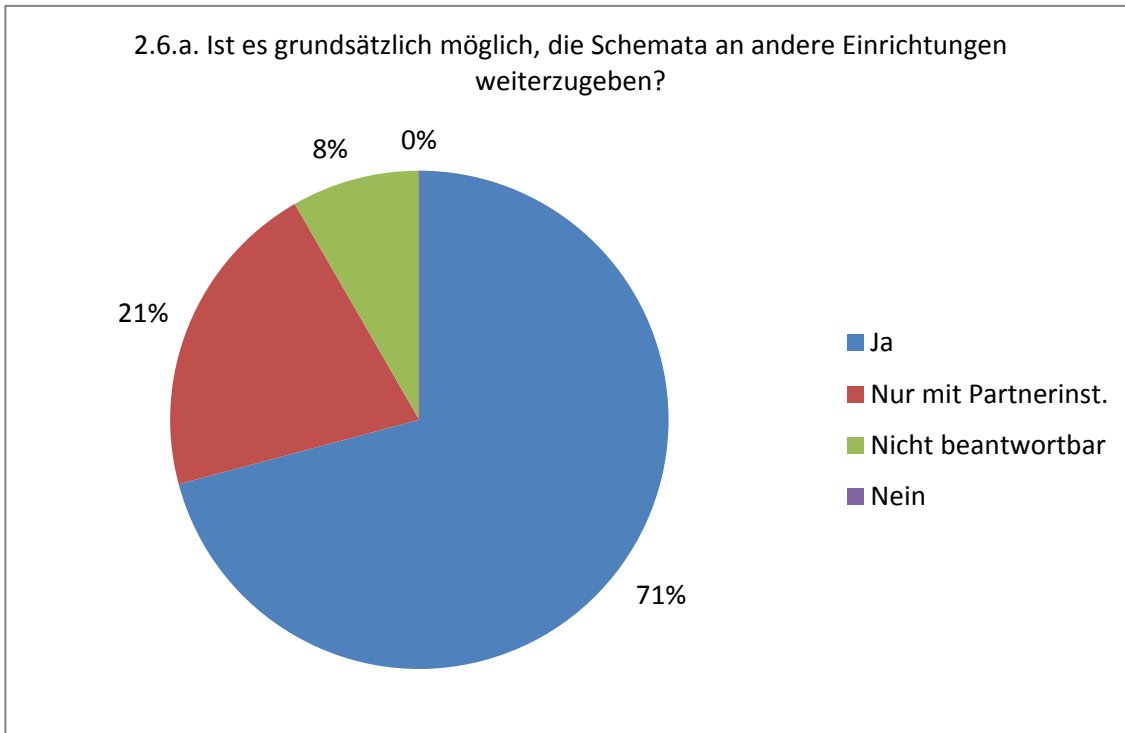
## 2.5. Wo sehen Sie die Grenzen der ausgewählten Schemata und Standards?

Die meisten InterviewpartnerInnen sahen (noch) keine Grenzen. Beim Beschreiben von sehr institutionsspezifischen Daten tritt gelegentlich der Wunsch nach zusätzlichen Metadatenfeldern auf. Andere Sonderwünsche wurden subsumiert unter dem Begriff „mangelnde Flexibilität“.



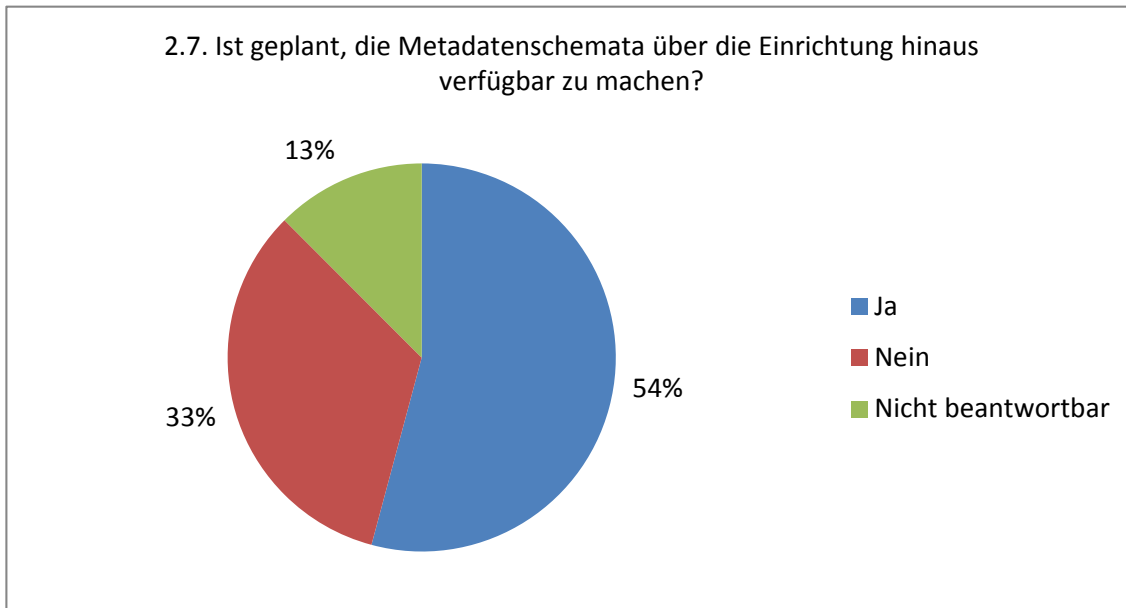
2.6. Ist es grundsätzlich möglich, die Schemata an andere Einrichtungen weiterzugeben? Wie wurde dies erreicht?

Bei dieser Frage waren sich bis auf jene Institutionen, die noch kein System haben und deshalb auch keine Antwort geben konnten, alle einig, dass die Schemata weitergegeben werden können. Abhängig vom gewählten System – wie z.B. bei Phaidra – geschieht dies durch den Austausch unter Partnerinstitutionen. Da fast alle ein offenes Format verwenden, ist eine prinzipielle Weitergabe zumeist von vornherein möglich.



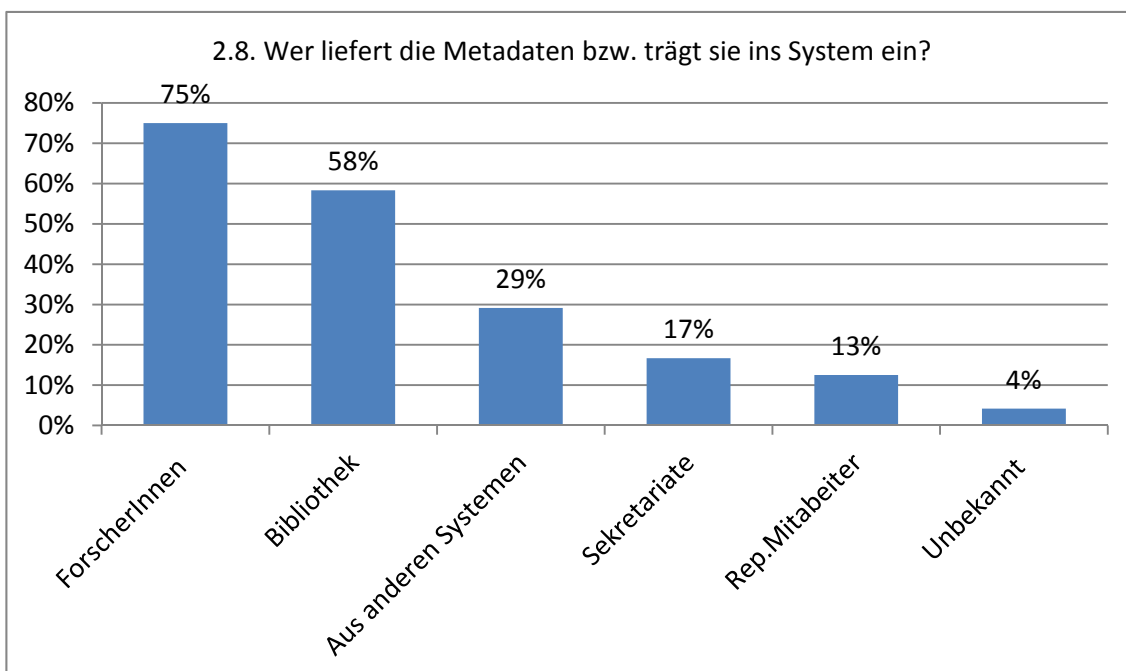
### 2.7. Ist geplant, die Metadaten schemata über die Einrichtung hinaus verfügbar zu machen?

Diese Frage wurde mehrheitlich mit „ja“ beantwortet, einige Institutionen hatten noch keine Entscheidung darüber getroffen.



### 2.8. Wer trägt die Metadaten ins System ein?

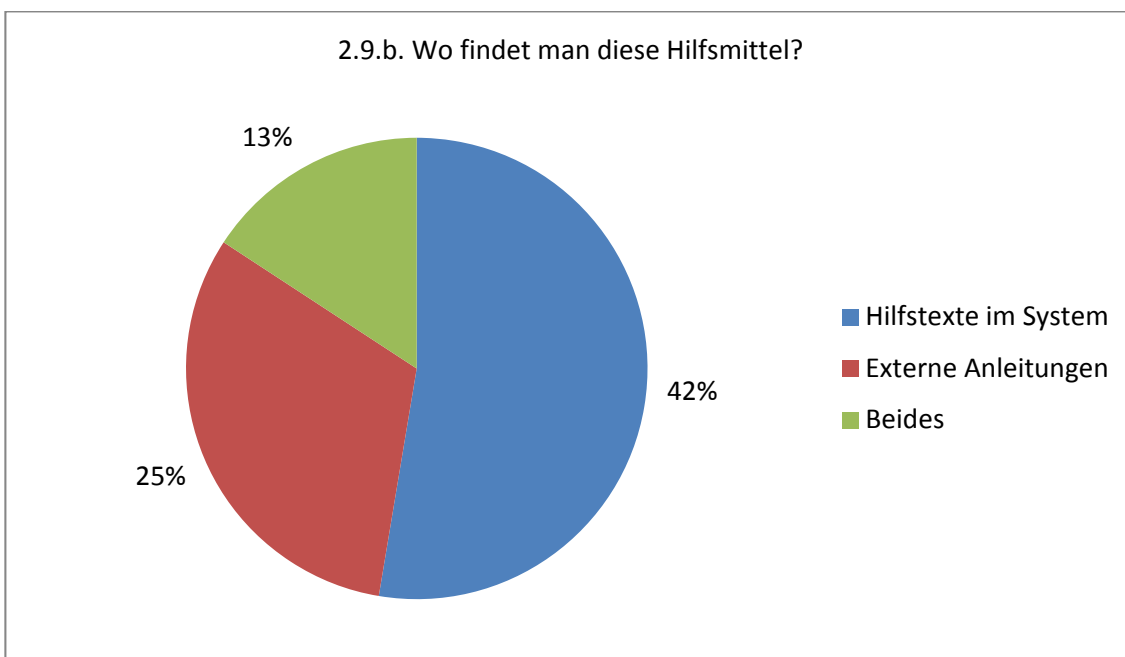
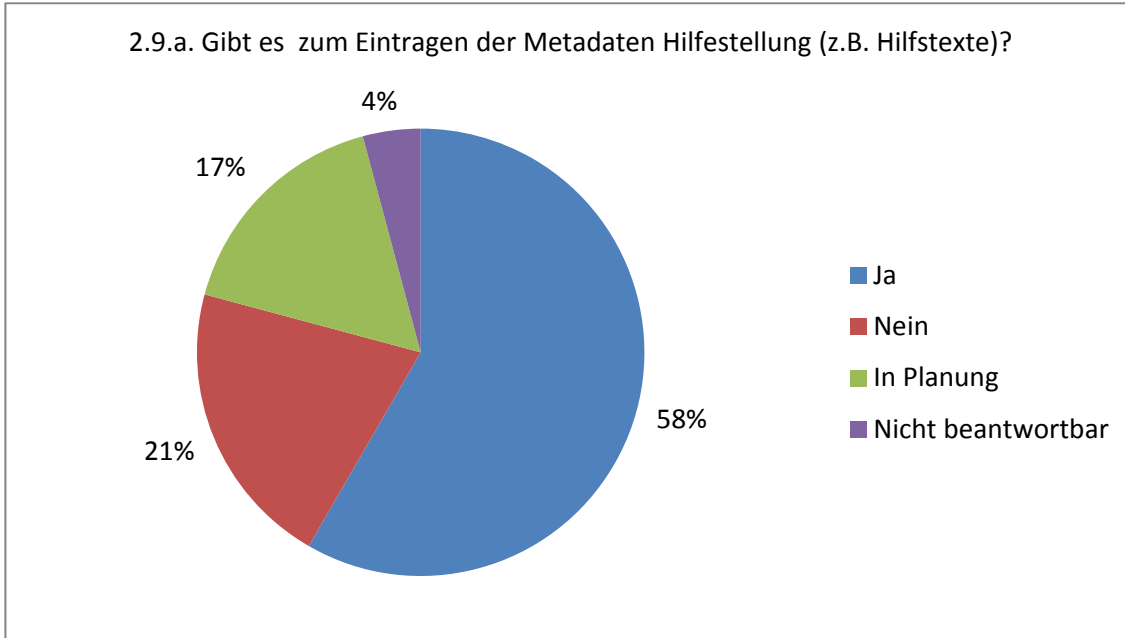
Interessant ist der hohe Anteil jener Einrichtungen, bei denen die ForscherInnen und WissenschaftlerInnen die Metadaten selbst eintragen. Gleich dahinter folgt das Modell, nach dem die Erfassung der Metadaten von der Bibliothek vorgenommen wird. In einigen Fällen werden die Metadaten (automatisiert) aus anderen Systemen übernommen, weniger oft durch die Verwaltungseinheiten auf Institutsebene (Sekretariate) oder speziell damit betraute RepositoriumsmitarbeiterInnen.





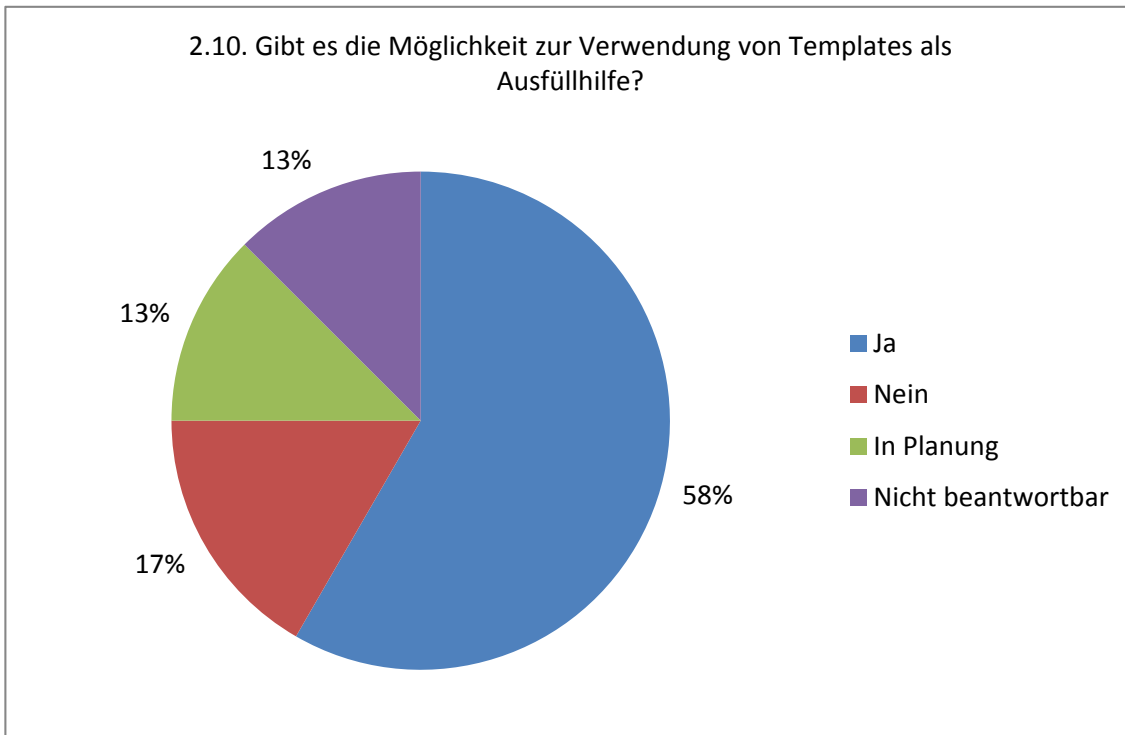
2.9. Gibt es zum Eintragen der Metadaten Hilfestellung (z.B. Hilfstexte, wie die Metadatenfelder auszufüllen sind)? Wo findet man diese?

Die überwiegende Mehrheit der Institutionen bietet Hilfestellungen an, und zwar entweder in Form von Hilfstexten, die direkt im System zu finden sind, oder als externe schriftliche Anleitungen. Viele Einrichtungen verfügen über beides.



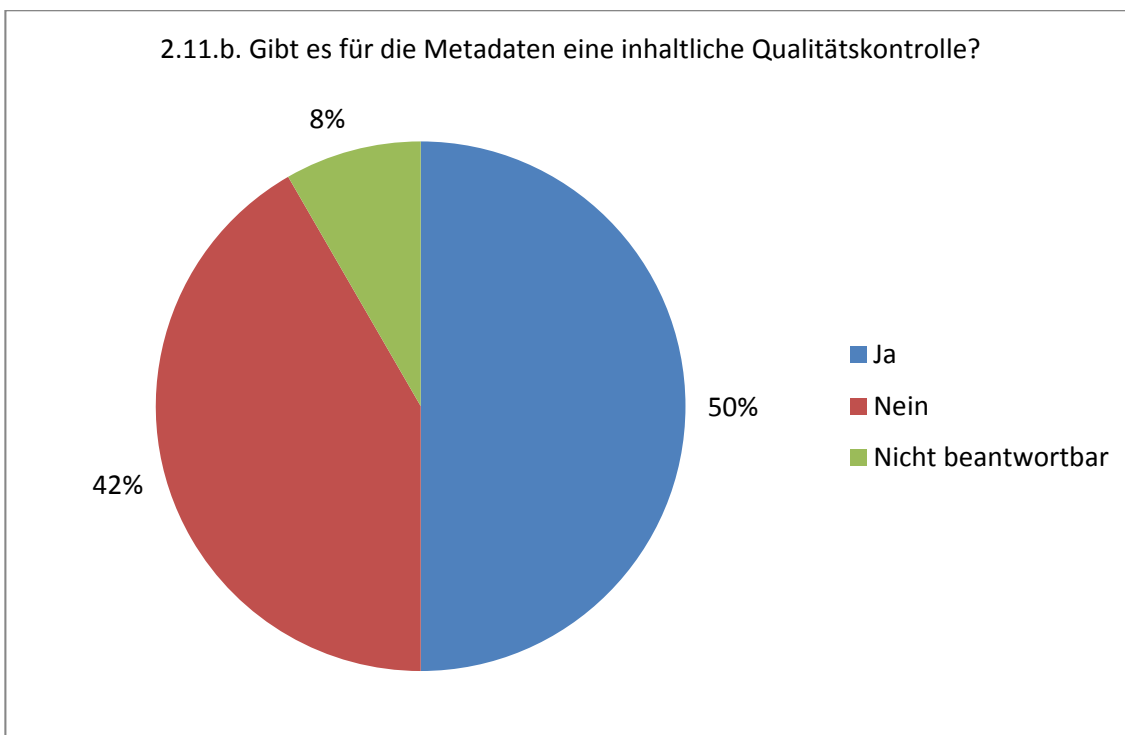
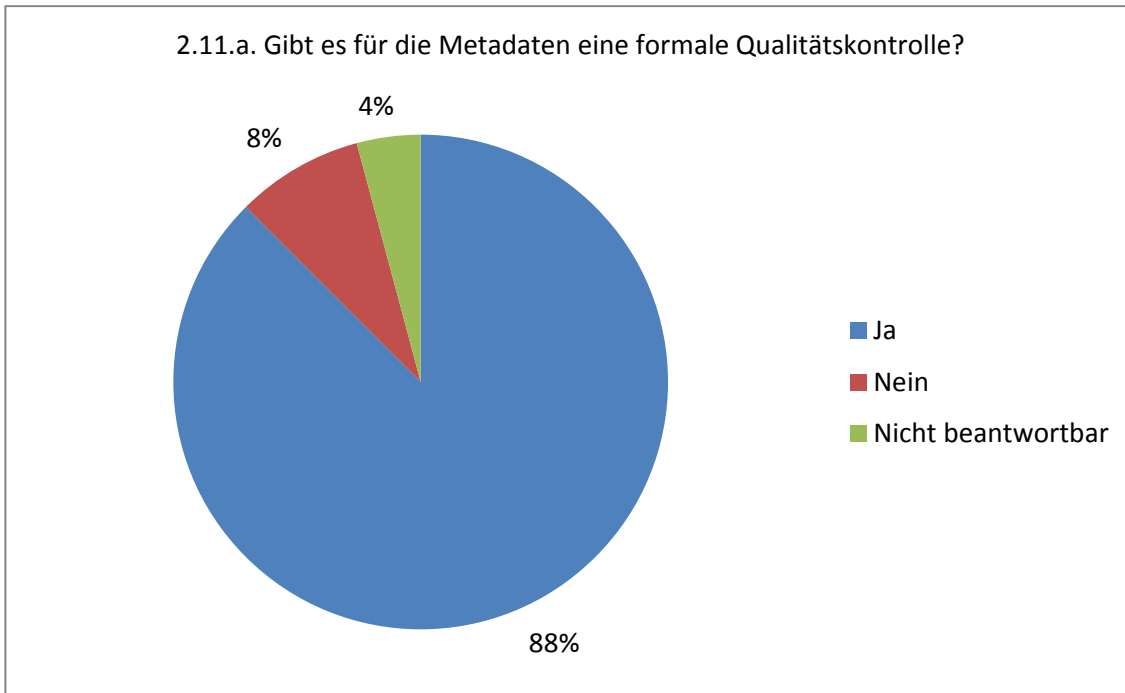
## 2.10. Gibt es die Möglichkeit zur Verwendung von Templates als Ausfüllhilfe?

Das Anbieten von Templates ist ebenfalls überwiegend bereits realisiert bzw. in Planung.



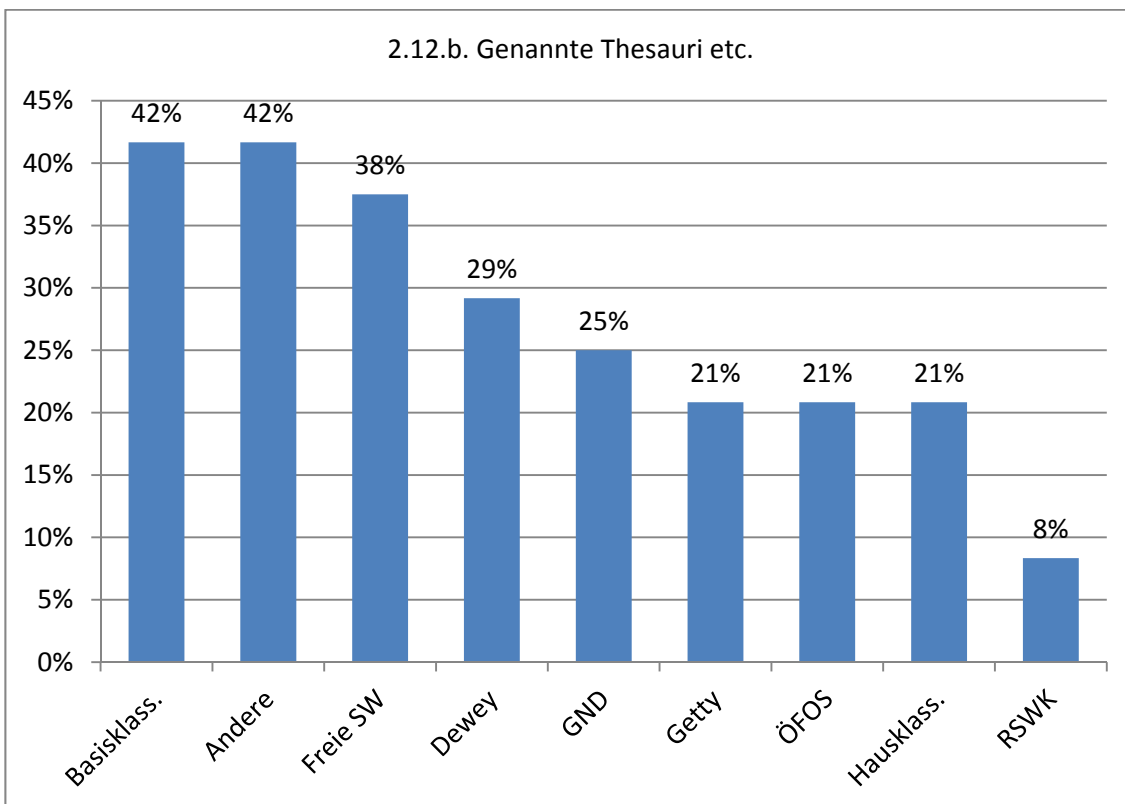
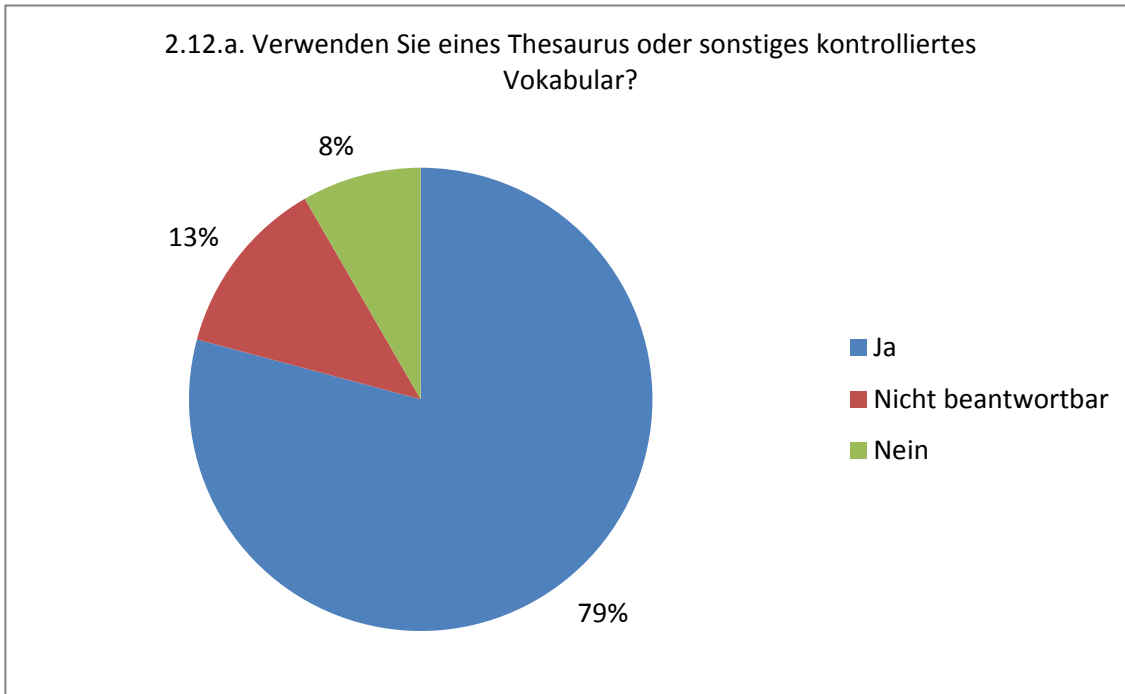
## 2.11. Gibt es für die Metadaten eine formale und / oder inhaltliche Qualitätskontrolle?

Nur wenige InterviewpartnerInnen gaben an, keine formale Qualitätskontrolle zu haben oder zu planen. Bei der inhaltlichen Qualitätskontrolle sieht es anders aus. Hier hat oder plant nur eine Minderheit der Institutionen keine oder nur eine sehr marginale Kontrolle. Viele konnten die Frage zu diesem Zeitpunkt aber auch noch nicht beantworten, da sie noch kein System hatten, bzw. es noch nicht im Regelbetrieb verankert war.



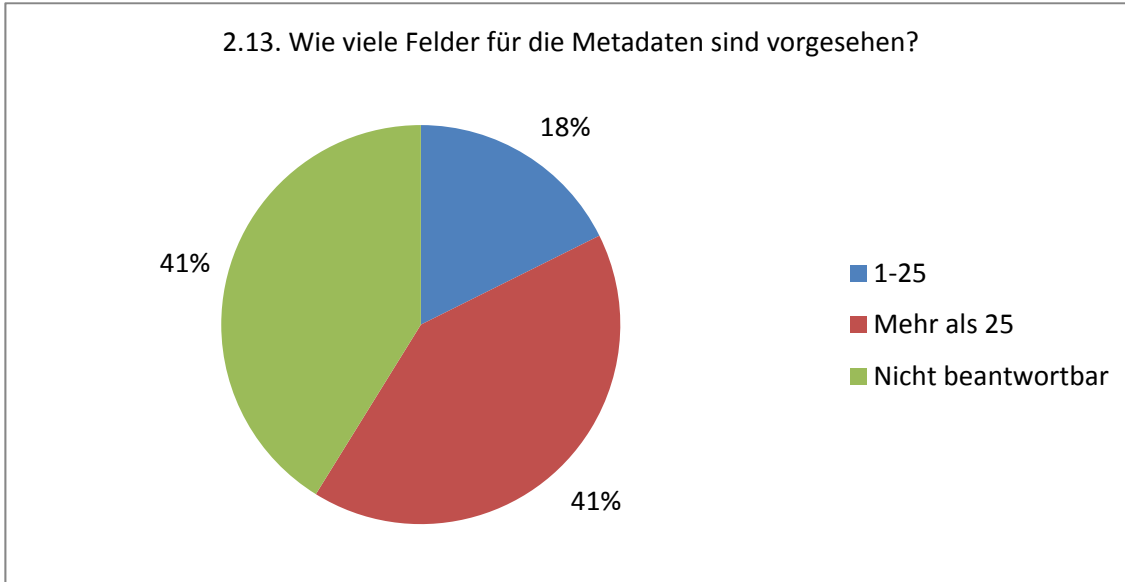
2.12. Verwenden Sie einen Thesaurus oder ein sonstiges kontrolliertes Vokabular? Wenn ja, welches?

Die überwiegende Mehrheit verwendet Thesauri oder ein kontrolliertes Vokabular. Die Grafik 2.12.b. zeigt, dass die Auswahl an Thesauri und Klassifikationen recht groß ist, einige nannten auch spezielle Thesauri, die hier unter „andere“ subsumiert sind.



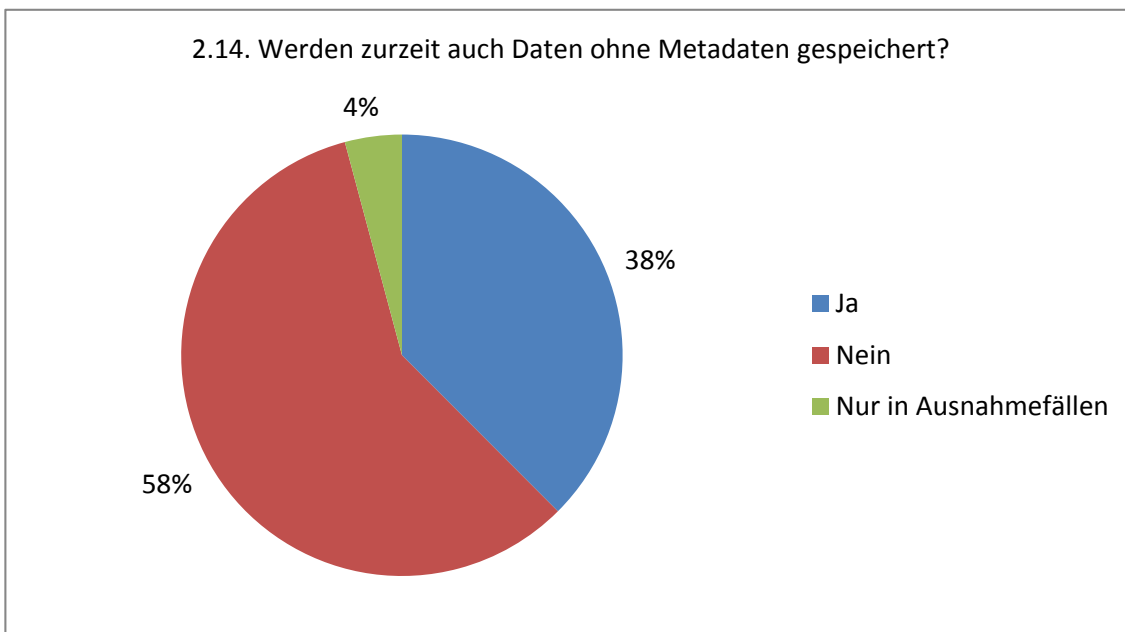
### 2.13. Wie viele Felder für die Metadaten sind vorgesehen?

Hier ist ein leichter Trend zu mehr als 25 Metadatenfelder zu erkennen. Einige InterviewpartnerInnen ergänzten jedoch, dass meist nicht alle dieser Felder verwendet werden. Der hohe Anteil von „nicht beantwortbar“ ergibt sich vor allem aus der Tatsache, dass einige der befragten Einrichtungen mehrere Repositorien (oder auch keines) betreiben und daher eine eindeutige Antwort nicht möglich ist.



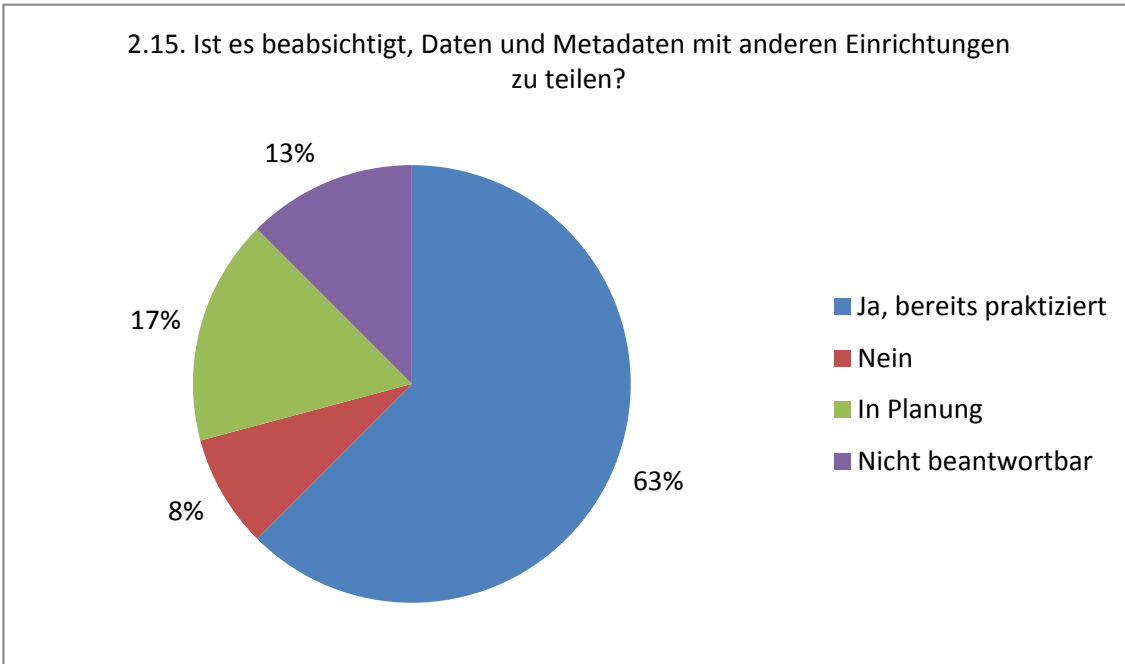
### 2.14. Werden zurzeit auch Daten ohne Metadaten gespeichert?

Daten ohne Metadaten werden zumeist außerhalb der offiziellen Repositorien (d.h. lokal auf Institutsebene etc.) gespeichert; die meisten Repositorien selbst lassen dies gar nicht zu. Die Antwort „Nein“ bedeutet i.d.R., dass über Praktiken der Datenerfassung außerhalb der Repositorien keine genauen Informationen vorliegen. Ziel ist in den allermeisten Fällen die Erfassung von Metadaten, die das Auffinden der beschriebenen Objekte erleichtern.



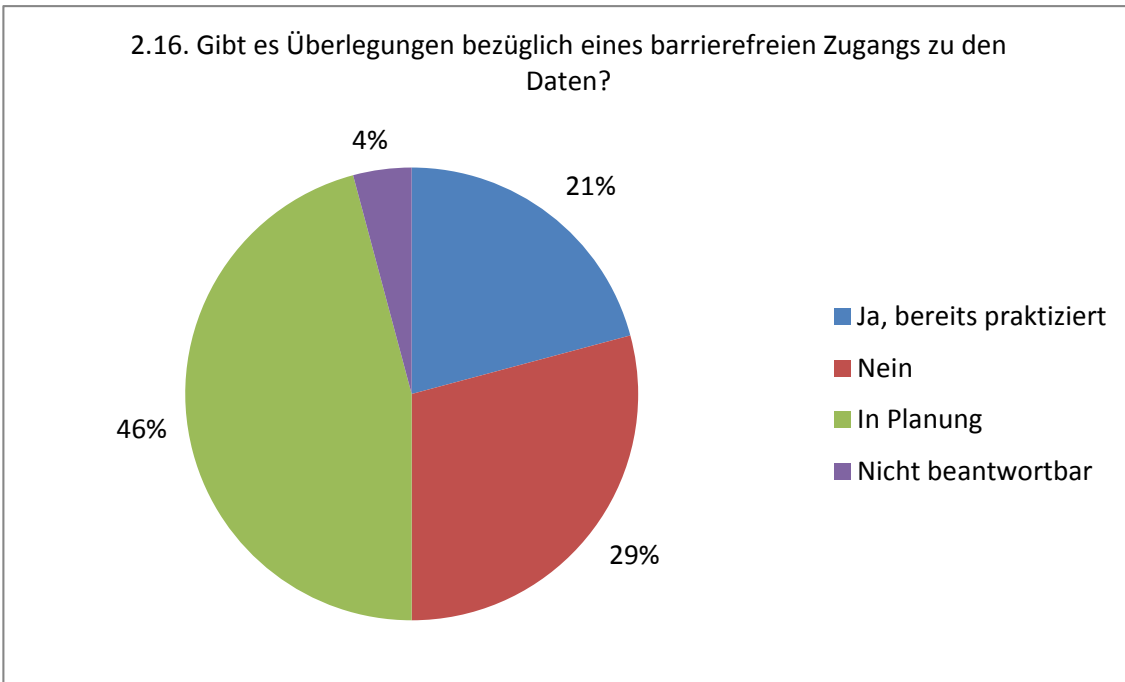
2.15. Ist es beabsichtigt, Daten und Metadaten mit anderen Einrichtungen zu teilen?

Eine deutliche Mehrheit der Befragten gibt an, die Daten und Metadaten mit anderen Einrichtungen teilen zu wollen und dies auch bereits zu praktizieren.



2.16. Gibt es einen barrierefreien Zugang zu den Daten?

Bemerkenswert ist, dass erst wenige Institutionen diese Frage bejahen konnten, was vorwiegend an den derzeit verwendeten, dafür nicht ausgelegten Systemen liegt. Die überwiegende Mehrheit strebt jedoch nach Barrierefreiheit und es gibt mehrere konkrete Projekte zur Umsetzung dieses Vorhabens.



## e-Infrastructures Austria

Sustainable data storage and the provision of data for use by third parties are the central roles of science. e-Infrastructures Austria is a federally funded program for the coordinated expansion and continued development of data repositories across Austria, and is made possible by a grant from the Austrian Ministry of Science, Research and Commerce (BMWFW). This program enables the safe archival and lasting availability of electronic publications, multimedia objects and other digital data from the research and teaching fields. Concurrently, topics relating to research data management and digital archiving workflows will be addressed.

### The working area is organized in twelve Work-Package-Clusters:

Cluster A	Monitoring of Document Repositories within the Partner Network <i>Patrick Danowski (IST Austria)</i>
Cluster B	Planning and Implementation of a „National Survey“ for Research Data <i>Christian Gumpenberger (University of Vienna)</i>
Cluster C	Designing a Knowledge Network: Development of a reference structure for the construction of Repositories <i>Paolo Budroni (University of Vienna)</i>
Cluster D	Infrastructure <i>Raman Ganguly (Vienna University Computer Center)</i>
Cluster E	Legal and Ethical Issues <i>Seyavash Amini (Counsellor-at-law, University of Vienna)</i>
Cluster F	Open Access <i>Andreas Ferus (academy of fine arts vienna)</i>
Cluster G	Visual Data modeling <i>Martin Gasteiner (University of Vienna)</i>
Cluster H	Life Cycle Management <i>Andreas Rauber (Technical University Vienna)</i>
Cluster I	Metadata <i>Susanne Blumesberger (University of Vienna)</i>
Cluster J	Permanent backup of the data <i>Adelheid Mayer (University of Vienna)</i>
Cluster K	Data from scientific and artistic-scientific research processes <i>Bernhard Haslhofer (Austrian Institute of Technology)</i>
Cluster L	Cross-project issues (technical and non-technical) <i>Andreas Jeitler (University of Klagenfurt)</i>