MASTER THESIS | MASTER‘S THESIS

Titel | Title

„Die Arbeit mit Metadatenformaten in digitalen Repositorien.
Eine Untersuchung anhand der Metadatenformate des digitalen Repositoriums PHAIDRA der Universität Wien.“

verfasst von | submitted by

Marija Ivanović BA MA

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (Library and Information Studies) (MSc)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 992 600
Universitätslehrgang lt. Studienblatt | Degree programme as it appears on the student record sheet:	Library and Information Studies
Betreut von | Supervisor:	Univ.-Prof. Mag. Dr. Gerhard Budin
Mitbetreut von | Co-Supervisor:	Mag. Sonja Fiala

Abstract auf Deutsch

Wissensorganisation ist ein zentrales Thema des Bibliotheks- und Informationswesens. Ein spezifischer Bereich ist dabei die Vergabe von Metadaten, um digitale Objekte in digitalen Repositorien zu beschreiben. Diese Arbeit untersucht, was Mappings zwischen unterschiedlichen Metadatenformaten in digitalen Repositorien notwendig machen kann und welche Vorteile sich durch Mappings ergeben. Ausgangspunkt der Analyse ist das digitale Repositorium PHAIDRA der Universität Wien. PHAIDRA steht für Permanent Hosting, Archiving and Indexing of Digital Resources and Assets. Um Zugang zum Wissen der PHAIDRA-Expert*innen zu gewinnen, werden teilstrukturierte Leitfaden-Interviews durchgeführt, die mittels der qualitativen Inhaltsanalyse mit deduktiver Kategorienanwendung ausgewertet werden.

Die Auswertung der Interviews zeigt, die Gründe, die Mappings zwischen unterschiedlichen Metadatenformaten notwendig machen sind zunächst die Notwendigkeit für eine Institution ein eigenes, komplexes, detailliertes und lokal entwickeltes Metadatenformat zu verwenden. Dies dient dazu den Bedarfen unterschiedlicher Nutzer*innen-Gruppen gerecht zu werden. Dies kann Mappings in andere Metadatenformate notwendig machen, um die eigenen Metadaten anderen Institutionen und Organisationen zugänglich zu machen oder in unterschiedlichsten Projekten mitwirken zu können und Interoperabilität herzustellen. Selbst wenn existierende Standards verwendet werden, können durch unterschiedliche Interpretationen des jeweiligen Standards Mappings zwischen den Interpretationen notwendig werden. Darüber hinaus können Projektpartner und Partnerorganisationen auch immer wieder die bei ihnen in Verwendung befindlichen Metadatenformat weiterentwickeln. Dies kann auch Mappings notwendig machen, um weiter kooperieren zu können. Auch Veränderungen am Datenmodell der Software, auf der das digitale Repositorium basiert, können Mappings notwendig machen.

Vorteile, die sich durch Mappings ergeben, sind bessere Möglichkeiten der Kooperation mit Partnern und in Projekten, wenn etwa gemeinsame Standards eingesetzt werden. Bei der spezifischen Änderung des Datenmodells ermöglicht dieses nun den Einsatz von global zugänglichen Standards, die persistente Identifikatoren einsetzen und damit langfristig nutzbar sind, während eine detaillierte Beschreibung beibehalten werden

kann. Außerdem fördert das neue Datenmodell auch Maschinenlesbarkeit und ermöglicht es Maschinen logische Schlüsse zu ziehen.

Abstract in English

Knowledge Organization is central to the field of Library and Information Studies. A specific area of application is the use of metadata for describing digital objects in digital repositories. This thesis analyzes, what makes mappings between different metadata formats necessary, and which advantages these mappings provide. Starting point of the analysis is the digital repository PHAIDRA of the University of Vienna. To gain access to the knowledge of the experts of PHAIDRA semi-structured expert interviews were conducted. They were analyzed using qualitative content analysis with deductive category assignment.

The analysis shows, the reasons which make metadata mappings necessary, begin with the use of local, complex and detailed metadata formats which are used to cater to the specific needs of the different stakeholders at the institution. A mapping becomes necessary to make the metadata available to other institutions or organization, to cooperate in different projects, and to enable interoperability. Even if organizations use community standards, a mapping might be needed as standards might be interpreted by different organizations in different ways. Additionally, project partners and partner organizations might further develop their own metadata formats, which again might call for a new or revised mapping, to be able to cooperate. Finally, changes in the software data model on which the repository is based, can call for a metadata mapping to the new structure.

The advantages mappings bring, are better cooperation with partners and in projects, e.g. when standards are used. Additionally, the specific change of the data model enables the use of globally accessible standards, which use persistent identifiers, and facilitate long term use, while permitting a detailed description of the digital object. Furthermore, the new data model is machine readable and enables machine reasoning.

Inhaltsverzeichnis

Abbildungsverzeichnis i
0. Einleitung 1
1. Forschungsstand 3
2. Wissensorganisation und Library and Information Studies 20
2.1. Der Zusammenhang zwischen den Bereichen Wissensorganisation und Library and Information Studies 20
2.2. Metadaten 22
2.3. PHAIDRA – das digitale Repositorium der Universität Wien 24
3. Prinzipien für gute Metadaten: die Prinzipien des „Framework of Guidance for Building Good Digital Collections“ und die FAIR-Prinzipien 26
3.1. Das “Framework of Guidance for Building Good Digital Collections” der National Information Standards Organization und seine Metadaten- 26
Prinzipien 26
3.2. Die FAIR-Prinzipien 29
3.3. Die NISO-Prinzipien für gute Metadaten und die FAIR-Prinzipien im Vergleich . 30
4. Die Methoden: das Expert*innen-Interview und die qualitative Inhaltsanalyse 37
4.1. Die Erhebungsmethode: das Expert*innen-Interview 37
4.1.1. Erstellung des Interviewleitfadens für das Expert*innen-Interview 38
4.1.2. Auswahl der Expert*innen/ Sampling 45
4.2. Die Auswertungsmethode: die qualitative Inhaltsanalyse 47
5. Auswertung der Expert*innen-Interviews 59
5.1. Verwendung von Community Standards 59
5.1.1. Auswahl des Schemas mit Bezug auf die in der Community verwendeten Schemata 59
5.1.2. Verwendung von Application Profiles 64
5.2. Unterstützung von Interoperabilität 65
5.2.1. Vergabe von detaillierten Metadaten als Grundlage für Interoperabilität 65
5.2.2. Zusammenarbeit zwischen Organisationen und Aggregatoren für Interoperabilität 70
5.2.3. Interoperabilität unterschiedlicher Metadaten-Schemata durch ihre Speicherung in Repositorien und generell öffentliche Dokumentation und Zugänglichmachung 71
5.2.4. Interoperabilität unterschiedlicher Metadaten-Schemata durch Mappings 72
5.2.5. Interoperabilität durch die Verwendung von Harvesting-Protokollen 74
5.3. Einsatz von standardisierten und kontrollierten Begriffen 75
5.3.1. Die Verwendung von standardisierten und kontrollierten Vokabularen, Thesauren und Taxonomien 75
5.3.2. Dokumentation der Verwendung externer als auch lokal entwickelter kontrollierter Vokabulare, Thesauren, Taxonomien 76
5.4. Festlegung der rechtlichen Nutzungsbedingungen 77
5.4.1. Beschreibung der Rechte 77
5.4.2. Management und Kommunikation von Rechten 78
5.4.3. Verwaltung rechtlich begründeter Zugangsbedingungen zu Objekten 78
5.5. Metadaten für langfristigen Erhalt und Pflege der digitalen Objekte 79
5.5.1. Der Einsatz von administrativen Metadaten 79

5.5.2. Der Einsatz von technischen Metadaten 80
5.5.3. Der Einsatz von Erhaltungsmetadaten 80
5.5.4. Der Einsatz von Recordkeeping-Metadaten für das Objekte-Management 81
5.5.5. Der Einsatz von strukturellen Metadaten 81
5.6. Metadaten als digitale Objekte – Einsatz eines persistenten, eindeutigen Identifikators 81
5.7. Die Modellierung von Metadaten: Metadatenkategorien und die Beziehungen zwischen ihnen 82
6. Diskussion 85
6.1. Diskussion zur Verwendung von Community Standards 85
6.2. Diskussion zur Unterstützung von Interoperabilität 87
6.3. Diskussion zum Einsatz von standardisierten und kontrollierten Begriffen 88
6.4. Diskussion zur Festlegung der rechtlichen Nutzungsbedingungen 89
6.5. Diskussion zur Verwendung von Metadaten für langfristigen Erhalt und Pflege der digitalen Objekte 90
6.6. Diskussion zur Verwendung von Metadaten als digitale Objekte – Einsatz eines persistenten, eindeutigen Identifikators 90
6.7. Diskussion zur Modellierung von Metadaten: Metadatenkategorien und die Beziehungen zwischen ihnen 90
7. Schlussbemerkungen 92
Anhang 94
Anhang 1: Interviewleitfaden für die Expert*innen-Interviews 95
Anhang 2: Interviewtranskripte 103
Anhang 3: E-Mail-Interviews 168
Anhang 3.1. E-Mail-Interview mit Susanne Blumesberger 168
Anhang 3.2. E-Mail-Interview mit Rastislav Hudak 171
Anhang 4: Kodierleitfaden mit Analysekategorien und Ankerbeispielen für die Auswertung der Interviews 177
Abkürzungsverzeichnis 196
Bibliographie 197

Abbildungsverzeichnis

Abbildung 1: Ablaufmodell der Auswertung: eine Kombination des allgemeinen Ablaufmodells (vgl. Mayring 2022:61) und des Ablaufmodells der deduktiven Kategorienanwendung (vgl. Mayring 2022:97). 55

0. Einleitung

Wissensorganisation ist wesentlich für die Strukturierung unterschiedlicher Inhalte im Bibliotheks- und Informationswesen. Ein Bereich, in dem ihre Prinzipien in Form von Wissensorganisationssystemen (engl. *knowledge organization systems*, kurz KOS) Anwendung finden, ist die Strukturierung von Metadaten in digitalen Repositorien. Diese Arbeit will anhand des Beispiels des digitalen Repositoriums PHAIDRA der Universität Wien untersuchen, was Mappings zwischen unterschiedlichen Metadatenformaten in digitalen Repositorien notwendig machen kann und welche Vorteile sich durch Mapping in andere Metadatenformate ergeben können.

Im ersten Kapitel dieser Arbeit findet sich der Forschungsstand rund um das Thema Mappings von KOS, der vier große thematische Felder umfasst. In einem ersten Themenblock werden Arbeiten zusammengetragen, die Unterschiede und Ähnlichkeiten von verschiedenen KOS untersuchen. Der zweite Themenbereich ergründet, welche methodischen Ansätze es für Mappings von KOS gibt. Der dritte thematische Bereich analysiert Studien zu bereits durchgeführten Metadaten-Mappings. Der vierte Bereich befasst sich mit den Ergebnissen eines Teils des Projekts e-Infrastructures Austria.

Das zweite Kapitel erarbeitet die Verbindung zwischen den Fachbereichen der Wissensorganisation und Library and Information Studies. Es wird der Begriff *Wissensorganisation* definiert, um dann zu zeigen, dass die Arbeit mit Metadaten zur Wissensorganisation gehört. Daraufhin wird dargestellt, welche Rolle Metadaten im Alltag haben, und welche Rolle sie über die Jahrhunderte für die Menschheit spielten ,wie sie in Bibliotheken, Archiven und Museen eingesetzt werden. Anschließend werden aktuelle Definitionen für den Begriff *Metadaten* betrachtet und das digitale Repositorium PHAIDRA der Universität Wien vorgestellt.

In Kapitel drei werden Prinzipien für gute Metadaten untersucht: Zunächst wird das *Framework of Guidance for Building Good Digital Collections* der *National Information Standards Organization* vorgestellt, das sechs Prinzipien für gute Metadaten bietet. Daraufhin werden die FAIR-Prinzipien für wissenschaftliche Daten und Metadaten

präsentiert, da sie gegenwärtig zentrale Empfehlungen für den Umgang mit wissenschaftlichen Daten und Metadaten bieten. Anschließend werden diese beiden Ansätze verglichen.

Kapitel vier beschreibt die in dieser Arbeit verwendeten methodischen Ansätze: Zunächst die Methode der Expert*innen-Interviews, die in der hier gewählten Form zur Informationsgewinnung dienen sollen. Dabei wird die Vorgehensweise zur Erstellung der Interviewleitfäden erarbeitet und die Auswahl der Expert*innen dargelegt. Anschließend wird die Auswertungsmethode der qualitativen Inhaltsanalyse und ihr Ansatz der deduktiven Kategorienanwendung nach Mayring (2022) vorgestellt, die als Analyseinstrumente für die im Rahmen dieser Arbeit durchgeführten Expert*innen-Interviews verwendet wird.

In Kapitel fünf findet schließlich die Auswertung der Expert*innen-Interviews statt. Dabei werden folgende Hauptkategorien sowie ihre jeweiligen Unterkategorien betrachtet: Verwendung von Community Standards, Unterstützung von Interoperabilität, der Einsatz von standardisierten und kontrollierten Begriffen, Festlegung der rechtlichen Nutzungsbedingungen, Einsatz von Metadaten für langfristigen Erhalt und Pflege der digitalen Objekte, Metadaten als digitale Objekte und die Modellierung von Metadaten.

Kapitel sechs ist der Diskussion der Ergebnisse im Hinblick auf die Forschungsfragen gewidmet und setzt diese mit dem Forschungsstand in Beziehung.

Das Kapitel sieben schließlich bietet eine Zusammenfassung der Ergebnisse.

.

1. Forschungsstand

Dieses Kapitel betrachtet bisherige Untersuchungen zu Mappings zwischen unterschiedlichen Arten von Wissensorganisationssystemen (engl. *knowledge organization systems*, kurz KOS), zu denen auch Mappings von Metadaten gehören. Die Untersuchungen können in vier große thematische Bereiche gegliedert werden: Es werden einerseits verschiedene Arten von KOS strukturell analysiert, dann werden mögliche methodische Ansätze für Interoperabilität, unter anderem auch durch Mappings, betrachtet. Schließlich gibt es auch Studien zu einzelnen konkret durchgeführten Mappings. Abschließend werden noch die Ergebnisse des Projekts e-Infrastructures Austria im Bereich des Clusters I: „Metadaten aus nicht-technischer und technischer Sicht" vorgestellt. Letzteres ist zwar keine klassische wissenschaftliche Literatur, doch das Projekt setzte sich eingehend mit der Arbeit mit Metadaten in Österreich auseinander und kann so ein Stück weit als Rahmen für die Analyse dieser Arbeit dienen. Die Untersuchungen zu diesen vier thematischen Bereichen sollen in diesem Kapitel vorgestellt werden.

- Strukturelle Untersuchungen zu unterschiedlichen Arten von KOS

Es gibt einige Arbeiten, die Merkmale unterschiedlicher Arten von KOS vergleichen, wie etwa Biagetti (2021), die Ontologien und unterschiedliche andere KOS, wie etwa Klassifikationssysteme oder Thesauri, analysiert und feststellt, dass Ontologien die „most evolved form of KOS" sind (Biagetti 2021:156), da sie eine Vielzahl von Beziehungen zwischen ihren Klassen darstellen können, was als „semantic richness" (Biagetti 2021:157) bezeichnet wird. Darüber hinaus versehen Ontologien ihre Klassen mit Attributen, was sie zusätzlich von anderen KOS unterscheidet. Biagetti (2021:158) verweist hier auf Zeng (2008).

Auch Marcondes (2013) untersucht die Ziele, die einerseits Ontologien und andererseits KOS inhärent sind. Dabei legt er im Bereich der traditionellem KOS einen Schwerpunkt auf Ranganathans (1967) Facettenklassifikation. Diese setzt Facetten als Komponenten ein, die kombiniert werden können und unterschiedliche Bereiche und Begriffe beschreiben. Sowohl Ontologien als auch KOS – und damit auch Ranganathans

Facettenklassifikation – strukturieren Wissen bestimmter Bereiche (engl. *domains*), jedoch haben Ontologien eine fundiertere ontologische (im klassischen Sinne) Basis, sind stärker logisch formalisiert und dadurch in der Lage, maschinelles logisches Schließen zu ermöglichen (vgl. Marcondes 2013:119). Deshalb sollte Ranganathans komplexer bibliothekarischer Ansatz der Facettenklassifikation mit den Prinzipien der zeitgenössischen Ontologien kombiniert werden, um in einer digitalen Welt zu bestehen, argumentiert Marcondes (vgl. Marcondes 2013:120). Für KOS im Allgemeinen, und die Facettenklassifikation im Speziellen, ist es daher notwendig, Prinzipien der Ontologie-Entwicklung zu berücksichtigen, um maschinelles Schließen zu ermöglichen (vgl. Marcondes 2013:120f.).

Diese beiden Artikel analysieren zwar unterschiedliche Arten von KOS (einschließlich der Ontologien), ihre Ziele und Funktionsweisen sowie theoretische Perspektiven. Sie bewegen sich jedoch auf einer strukturellen Metaebenen, wägen dabei unterschiedliche Aspekte ab und gehen noch nicht im Detail auf die Ursachen ein, die ein Mapping zwischen zwei KOS notwendig machen könnten.

Lauruhn und Groth (2016) beleuchten die Gründe für Veränderungen der internen Strukturen und Elemente von unterschiedlichen KOS und stellen neun Gründe fest: „changing norms, political influence, new developments, maintenance; incremental contributorship; progressive formalization; software and automation; integration of large numbers of data sources; and variance in algorithm training data“ (Lauruhn & Groth 2016:628). Die ersten drei treten typischerweise bei traditionsreichen KOS wie der Dewey Decimal Classification (DDC) oder den Library of Congress Subject Headings (LCSH) auf. Die letzten sechs Gründe entstehen durch neue Ansätze der KOS-Entwicklung, insbesondere durch Automatisierung und Crowd-Sourcing-Ansätze (vgl. Lauruhn & Groth 2016:628). Diese unterschiedlichen Gründe für die Veränderung der inneren Strukturen von KOS werden langfristig auch eine Veränderung der klassischen Aufgaben von Expert*innen, die KOS entwickeln, zur Folge haben. Dabei werden Software und Algorithmen Aufgaben wie das Entdecken und die Empfehlung von neuen Begriffen (engl. *concepts*) übernehmen, während die Expert*innen diese dann analysieren, ihren Ursprung bewerten und sie gegebenenfalls in das jeweilige KOS einarbeiten. Auf dieselbe Art könnten auch Vorschläge, die durch

Crowd-Sourcing entstehen, bewertet, analysiert und in ein KOS integriert werden (vgl. Lauruhn & Groth 2016:627). Die Arbeit von Lauruhn und Groth (2016) zeigt, welche Gründe zu interner Umstrukturierung von KOS führen können und deutet auf eine Weiterentwicklung des Aufgabenprofils von Expert*innen in diesem Bereich hin. Allerdings wird nicht beleuchtet, ob diese Aspekte oder andere Gründe ein Mapping von einem KOS in ein zweites KOS notwendig machen würden.

Die Veränderungen, die innerhalb von KOS notwendig werden können, zeigt auch Fiala (2019) für die Darstellung von barrierefreien Dateien auf. Sie analysiert dafür – als Beispiel für das in unterschiedlichen Repositorien noch bestehende Problem[1] – die unterschiedlichen Möglichkeiten der Kennzeichnung in MARC21 und MODS (kurz für Metadata Object Description Schema) von barrierefreien Dateien. Dabei sollen neue Inhalte dargestellt werden, die in die bestehenden Strukturen eingearbeitet werden müssen. Diese neuen Aspekte müssen dann auch die Möglichkeit bieten, in weitere Metadatenformate gemappt zu werden. Ob die Notwendigkeit, neue Informationen in bestehende Strukturen einzubetten, Mappings von Metadatenformaten generell erforderlich machen kann, wird jedoch nicht ausgeführt.

- Untersuchungen zu methodischen Ansätzen für Interoperabilität

Mit methodischen Ansätzen für Interoperabilität, und dazu gehören auch Mappings, befassen sich etwa de Almeida Campos, Campos und Barbosa (2020): Sie untersuchen, wie die Bereiche der Informationswissenschaft und der Informatik Fragen rund um die semantische Interoperabilität unterschiedlicher Datenbanken lösen können (vgl. de Almeida Campos u. a. 2020:683f.). Dafür analysieren sie die klassischen Ansätze der Informationswissenschaft von Neville (1970; 1972), der eine „intermediate language“, entwickelte und Dahlberg (1981a; 1981b; 1983), die eine Kompatibilitätsmatrix für Begriffe erarbeitetet. Beide Ansätze versuchen Interoperabilität zwischen KOS zu ermöglichen. Diese Ansätze erfordern jedoch intellektuelle und manuelle Arbeit in einem großen Maß und sind daher nicht in der Lage, heutige Datenmengen zu bewältigen (vgl. de Almeida

[1] Das Repositorium EMMA (Educational Materials Made Accessible) hat sich dagegen auf die Zugänglichmachung und Beschreibung von barrierefreien Dateien spezialisiert und stellt diese den an EMMA teilnehmenden Institutionen und Repositorien zur Verfügung (vgl. EMMA 2024:2).

Campos u. a. 2020:689f.). Dafür müssten Techniken verwendet werden, die im Umfeld der und durch die Informatik entwickelt werden: Programme, die automatisiert arbeiten und unterschiedliche Datenbestände mit Rücksicht auf die Übertragung der Bedeutungen ihrer Elemente mappen und integrieren können. Zusätzliche Vokabulare zu entwickeln ist dabei nicht das Ziel, vielmehr sollte eine Art vermittelnder Zwischensprache als Referenzpunkt zwischen unterschiedlichen Vokabularen entwickelt werden, die es ermöglicht, diese nicht ausschließlich durch ihre Zeichenfolge, sondern auch semantisch zu erfassen (vgl. de Almeida Campos u. a. 2020:690ff.).

Für Kempf et al. (2014:67) sind Mappings geprägt durch zeitintensive und teure Arbeit, die durch Automatisierungsprozesse erleichtert werden könnte. In ihrer Untersuchung wenden sie 21 Ontology-Matching-Tools auf die Thesauri *Thesaurus for the Social Sciences* und den *Thesaurus for Economics* an und vergleichen einerseits die Resultate des automatisierten Crosswalks, wie das Mapping zwischen hier zwei Thesauri auch bezeichnet werden kann, mit einem vorher (vgl. Kempf u. a. 2014:69) manuell durchgeführten Crosswalk derselben Thesauri, der von Mayr und Petras (2008) durchgeführt wurde (vgl. Kempf u. a. 2014:68). Das Mapping schlossen 13 der 21 Tools ab (vgl. Kempf u. a. 2014:69). Die Korrektheit der Mappings lag dabei zwischen 6% und gerundet 54% richtiger Zuordnungen der jeweiligen Begriffe der beiden Thesauri. Ein ausschließlich automatisch durchgeführtes Mapping würde damit keine zufriedenstellenden Ergebnisse liefern (vgl. Kempf u. a. 2014:70). Durch die Empfehlungen der Ontology-Matching-Tools und die Reorganisation und das In-Beziehung-Setzen ihre Resultate, kann jedoch die zeitintensive intellektuelle Arbeit von Expert*innen rund um Mapping-Prozesse beschleunigt werden (vgl. Kempf u. a. 2014:73).

Für Chan (2005) wollen Nutzer*innen mit einer Suche möglichst viele digitale Bestände von unterschiedlichen Bibliotheken und Archiven durchsuchen können. Um dies zu erreichen, müssen die zur Beschreibung dieser Ressourcen verwendeten Metadaten-Schemata interoperabel sein. Grundlage für Interoperabilität ist die Möglichkeit, Daten auszutauschen und von einem Format in ein anderes umzuwandeln, also zu mappen. Chan beleuchtet sieben Ansätze, die Interoperabilität ermöglichen sollen: Den Einsatz eines einheitlichen Metadaten-Standards, die Modifikation/Adaption eines bestehenden

Schemas, um den Bedürfnissen einer bestimmten Nutzer*innen-Gruppe gerecht zu werden. Dies schließt auch die Methode des Application Profiling ein, das definiert, in welcher Form ein bestimmtes Metadaten-Schema von einer Nutzer*innen-Gruppe angewandt werden soll. Eine weitere Methode ist die Ableitung (engl. *derivation*) eines einfacheren Schemas von einem bestehenden komplexeren Schema. Darüber hinaus gibt es noch Crosswalks, die Äquivalenzen zwischen den einzelnen Elementen unterschiedlicher Schemata suchen. Meist geschieht ein Crosswalk zwischen zwei Schemata (in eine oder beide Richtungen), da sich beim Mapping zwischen mehreren Schemata, die Zahl der notwendigen Mappings vervielfacht. Ein Ansatz, der Mappings zwischen mehreren Schemata möglich macht, ist das Switching: dabei fungiert ein Schema als Referenzpunkt, über den die Mappings durchgeführt werden. Falls es kein geeignetes Referenz-Schema geben sollte, kann auch mit der Lingua Franca als Methode gearbeitet werden. Diese versammelt die häufigsten gemeinsamen Elemente der verwendeten Schemata in einer übergeordneten Struktur, ist selbst aber kein Schema. Abschließend besteht noch die Möglichkeit ein Metadaten-Framework wie etwa RDF[2] als eine Art Container zu verwenden, der es ermöglicht, unterschiedliche Standards in einer Darstellungsform zu umfassen, indem die Elemente der unterschiedlichen Metadaten-Schemata integriert werden. Diese Methoden sollen digitale Ressourcen optimal zugänglich machen (vgl. Chan 2005).

Auch Chan und Zeng (2006a; 2006b) sehen wie Chan (2005) die Notwendigkeit, eine Suche über mehrere Repositorien zu ermöglichen und dafür Metadaten von Repositorien unterschiedlicher Bereiche abzufragen. Sie untersuchen daher Methoden, die verwendet werden, um Interoperabilität zwischen unterschiedlichen Metadaten-Schemata und Anwendungsszenarien zu erreichen. Dabei analysieren sie zunächst die Szenarien, die den Einsatz und die Interoperabilität von Metadaten in einzelnen Projekten oder in integrierten Repositorien notwendig machen. Dazu gehört die Verwendung von individuellen Metadaten-Schemata in einzelnen Projekten, dann der Einsatz von Application Profiles, die Elemente unterschiedlicher Schemata auswählen und zusammenführen. Des

[2] Das Resource Description Framework (kurz: RDF) ist ein Standard des World Wide Web Consortiums (W3C), der es ermöglicht Daten im Internet auszutauschen. Durch die Verwendung von RDF können Daten zusammengeführt werden, selbst wenn sie auf unterschiedlichen Schemata basieren. Außerdem unterstützt RDF auch die Weiterentwicklung von Schemata, ohne dass es dabei zu Änderungen für die Nutzer*innen (sei es Mensch oder Maschine) kommen muss (vgl. W3C 2024).

Weiteren gehören dazu der Austausch von Metadaten-Datensätzen zwischen Datenbanken auf Grundlage von sich entsprechenden Elementen der jeweiligen Schemata und abschließend das Einsammeln (engl. *harvesting*) oder Verknüpfen (engl. *merging*) von unterschiedlichen Metadatensätzen durch ein oder in einem Repositorium. Diese Perspektive auf die Szenarien wird von Chan und Zeng (vgl. 2006a) durch den Blick auf drei Ebenen ergänzt, die Interoperabilität ermöglichen: die Ebene des Schemas, die Ebene des jeweiligen Eintrags (engl. *record*) und die Ebene des Repositoriums. Chan und Zeng (vgl. 2006a) analysieren zunächst, welche Methoden Interoperabilität auf Schema-Ebene ermöglichen. Dabei führen sie, wie Chan (2005) auch, Ableitung und die Erstellung von Application Profiles an, sowie Crosswalks und Switching. Chan und Zeng (vgl. 2006a) fügen, Chan (2005) ergänzend, den Einsatz von Metadata-Registries hinzu, die Informationen über unterschiedliche Metadaten-Schemata sammeln und sie unabhängig vom Fachbereich, zu dem sie gehören, in einer Datenbank zugänglich machen. Im zweiten Teil ihre Untersuchungen betrachten Chan und Zeng (vgl. 2006b) die Methoden, die auf Eintrags-Ebene eingesetzt werden können, um Interoperabilität zu erreichen, sowie die Methoden, die auf Repositoriums-Ebene eingesetzt werden können. Auf Eintrags-Ebene sind das die Umwandlung von Metadaten-Datensätzen: dabei werden bestehende Datensätze, die in einem Metadaten-Schema vorliegen, in Datensätze eines anderen Metadaten-Schemas übersetzt. In diesem Prozess werden die Inhalte einzelner Datenkategorien/ Elemente eines Schemas in (möglichst) äquivalente Datenkategorien/ Elemente des anderen Schemas übertragen, also gemappt. Dieser Prozess ist auf gewisse Weise auch ein Crosswalk, wie er auf Schema-Ebene beschrieben wurde, nur werden hier komplette Datensätze, also auch die Inhalte der Datenkategorien konvertiert. Die zweite Möglichkeit zur Schaffung von Interoperabilität auf Datensatz-Ebene ist Integration. Dabei können durch einen modularen Ansatz Elemente etwa unterschiedlicher Metadaten-Schemata und -Vokabulare kombiniert werden, um Interoperabilität zu erreichen. Dazu gehört auch der Einsatz von Containern, etwa von RDF (vgl. Chan & Zeng 2006b), wie auch von Chan (2005) beschrieben.

Abschließend analysieren sie, wie auf Ebene von Repositorien Interoperabilität erreicht werden kann: Dafür kann Metadaten-Harvesting, also das Sammeln von Metadaten aus unterschiedlichen Quellen in einem Metadaten-Repositorium angewandt werden,

wie etwa im Fall des National Science Digital Library Metadata Repository. Dieses ermöglicht zusätzlich die Eingabe von Metadaten mit möglichst geringem menschlichem Aufwand, mittels des Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), das standardisiert, wie die Daten zwischen Repositorien ausgetauscht werden können. Daneben gibt es Metadaten-Repositorien, die unterschiedliche Metadatenformate unterstützen, ohne diese umzuwandeln: So etwa das Digital Library for Earth System Education Collections System, das es ermöglicht Sammlungen von Metadaten-Datensätzen, die unterschiedliche Metadaten-Schemata in XML[3] verwenden können, zu erstellen und diese weiterzuentwickeln, zu durchsuchen und zu teilen, ohne die einzelnen Datensets transformieren zu müssen. Dafür wird gegebenenfalls auch auf die jeweils verwendeten Schemata verwiesen (vgl. Chan & Zeng 2006b).

Auch Aggregation, das Sammeln von sich ergänzenden Metadaten (teilweise unterschiedlicher Qualität) aus unterschiedlichen Quellen ist eine Möglichkeit, um auf Repositorium-Ebene Interoperabilität herzustellen. Teilweise müssten die Metadaten in diesem Fall auch überarbeitet werden, um vollständigere Metadaten zu erhalten, (vgl. Chan & Zeng 2006b). Dann gibt es noch die Möglichkeit Crosswalk-Services einzusetzen, um auf Schema-Ebene die Elemente oder auf Werte-Ebene die Werte, die ein Element in einem bestimmten Datensatz angenommen hat, zu mappen. Wenn etwa in einem Repositorium unterschiedliche Metadaten aus unterschiedlichen Quellen gesammelt werden (vgl. Chan & Zeng 2006b). Außerdem können Schlagwörter aus (auch sprachlich) verschiedenen Schlagwortkatalogen gemappt und so zueinander in Beziehung gesetzt werden. Bei einer Suche über mehrere Metadaten-Repositorien hinweg eingesetzt, können so Ressourcen mit gleichen Schlagwörtern in mehreren Sprachen gefunden werden (vgl. Chan & Zeng 2006b). Zusätzlich können die Schlagwörter, die einer Ressource bei der Aufnahme unter Verwendung mehrerer kontrollierter Vokabulare zugeordnet werden, als gleichwertig betrachtet werden. Auf diese Weise kann automatisch ein loses Mapping zwischen mehreren kontrollierten Vokabularen hergestellt werden (vgl. Chan & Zeng 2006b).

[3] XML (Extensible Markup Language) ist eine textbasierte Auszeichnungssprache, die dazu dient, strukturierte Daten darzustellen. Sie kann sowohl von Menschen als auch Maschinen gelesen werden und ermöglicht auf diese Art das Teilen von strukturierten Daten (vgl. W3C 2023).

Die hier versammelten Arbeiten befassen sich damit, wie Interoperabilität zwischen Datensätzen mittels unterschiedlicher methodischer Ansätze erreicht werden kann. Dabei wird die Notwendigkeit der Interoperabilität, um Metadaten auffindbar zu machen erwähnt. Jedoch wird nicht untersucht, welche Ursachen, im Hinblick auf die vorhandenen Metadaten(-Strukturen), Mappings notwendig machen.

- Beschreibung unterschiedlicher Mapping-Prozesse

Darüber hinaus gibt es Beschreibungen von konkreten Mapping-Prozessen, wie etwa bei Barroso, Hartmann und Ribeiro (2015), Southwick (2015), Bellotto und Bettella (2019) und Fiala und Huggle (2019). Barroso, Hartmann und Ribeiro (2015:36) beschreiben den Prozess, der an der Universität Porto eingesetzt wurde, um Interoperabilität zwischen den Sammlungen des Archivs, der Bibliothek und des Museums der Faculdade de Belas Artes der Universisdade de Porto herzustellen. Dies war wichtig, da die unterschiedlichen Sammlungen inhaltlich eng verknüpft sind: Sie enthalten alle Werke bedeutender Maler, wie etwa von António Carvalho de Silva Porto. Jedoch war eine Suche über alle drei Sammlungen von einem Ausgangspunkt bisher nicht möglich. Die Sammlungen der Bibliothek und des Archivs waren bereits in der Biblioteca Digital de Arte (BDArt – Arts Digital Library) vorhanden, die im Thematischen Repositorium der Universität Porto (*pt. Repositório Temático da Universidade de Porto*) gehostet wird. Um die Sammlung des Museums vom gleichen Suchpunkt aus zugänglich zu machen, wurde ein Crosswalk zwischen dem ursprünglichen Metadatenformat der Museumssammlung, ICOM/CIDOC[4] und dem Zielformat Dublin Core (DC) entwickelt (vgl. Barroso u. a. 2015:37). Dafür musste zunächst die Festlegung der relevantesten Metadatenfelder im Metadatenformat der Museumssammlung stattfinden, da dieses Ausgangsformat über 300 Metadatenfelder verfügte. Zusätzlich musste die Auswahl der Metadatenfelder mit den in der BDArt bereits verwendeten Metadatenfeldern abgestimmt und harmonisiert werden. Die gewählten Metadatenfelder der Museumssammlung wurden dann im Hinblick auf ihre tatsächliche Nutzungsart untersucht, bevor schließlich die entsprechenden Metadatenfelder in Dublin

[4] Dieses Metadatenformat basiert auf den Empfehlungen des International Council of Museums/ International Committee for Documentation.

Core ausgewählt wurden, die bereits im Thematischen Repositorium der Universität Porto verwendet wurden (vgl. Barroso u. a. 2015:38ff.). Nachdem der Crosswalk definiert worden war, wurden die Metadatenfelder in einem XML-Format exportiert und in das Repositorium importiert (vgl. Barroso u. a. 2015:41). Dadurch ist es möglich geworden, die Inhalte der unterschiedlichen, aber inhaltlich verwandten Sammlungen von einem einzigen Einstiegspunkt zu durchsuchen (vgl. Barroso u. a. 2015:43).

Southwick (2015) beschreibt detailliert das Vorgehen an der Universität Nevada in Las Vegas (UNVL), als diese die Metadaten der digitalen Sammlung der UNVL in Linked Open Data (LOD) umwandelte. Wobei diese Beschreibung auch anderen Institutionen Ansatzpunkte für die Transformation von Metadaten in LOD bieten könnte (vgl. Southwick 2015:34). Zu diesem Prozess gehörte auch ein Mapping des bisher verwendeten Datenmodells der Metadaten der digitalen Sammlung in das zukünftig zu verwendende Datenmodell, in diesem Fall das Europeana-Datenmodell, und die Bereinigung der Daten (vgl. Southwick 2015:16f.). Motiviert war die Transformation der Metadaten in LOD durch drei Faktoren: zunächst die Möglichkeit, die bisher schwer zugänglichen Metadaten der digitalen Sammlungen einem größeren Forum von Nutzer*innen zugänglich zu machen; dann die eigenen Daten mit den Daten anderer Institutionen und Organisationen zu verbinden. Dies würde drittens die hürdenlose Suche für die Nutzer*innen ermöglichen, unabhängig davon, wo die Daten erstellt worden waren (vgl. Southwick 2015:2). Damit geht der Ansatz von Southwick noch über die Interoperabilität von verschiedenen Sammlungen innerhalb einer Institution hinaus und ermöglicht den Zugang zu Objekten für unterschiedlichste Nutzer*innen-Gruppen.

Bellotto und Bettella (2019) betrachten den Einsatz des Datenmodells von PHAIDRA, das an der Universität Wien entwickelt wurden, an der Universität Padua. Bellotto und Bettella analysieren darin das Mapping von UWMETADATA nach Dublin Core, das für den OAI-PMH Data Provider von PHAIDRA eingesetzt werden sollte, und das Mapping nach MODS. Zusätzlich betrachten sie die Darstellung der Metadaten in PHAIDRA und die notwendigen Schritte für eine Entwicklung in Richtung Semantic Web (vgl. 179ff. Bellotto & Bettella 2019). Bei der Untersuchung des Mapping von UWMETADATA nach Dublin Core notieren sie sehr kurz einige Probleme und Vorteile in Zusammenhang mit diesem Mapping. Dazu gehört, dass dieses Mapping ein besseres

Verständnis für die technischen Ansprüche, die ein komplexeres Metadatenformat, wie etwa MODS hat, bietet. Außerdem nennen sie die verbesserten Entscheidungen im Hinblick auf die Arbeit mit Metadatenelementen, die durch das Mapping von UWMETADATA auf MODS für das Mapping nach Dublin Core getroffen werden konnten. Auch wurde durch die Analyse des Formats UWMETADATA und dessen stärkerer Formalisierung eine präzisere Darstellung des Formats, seiner Daten und Inhalte ermöglicht (vgl. Bellotto & Bettella 2019:177). Darüber hinaus machte das Mapping die zentrale Bedeutung von Standards für die Interoperabilität, die Zugänglichkeit und die Wiederverwendung in anderen Kontexten deutlich. Es wurde dadurch zusätzlich ermöglicht, die Methodik des Daten-Lesens und Verstehens als Prozess zu begreifen, der zwischen vielfältigen Schemata, Strukturen, ihren Manifestationen, Modellen und Korpora vermittelt, was immer wieder Verständnis für diese Elemente und die Arbeit mit Kompromissen fordert (vgl. Bellotto & Bettella 2019:177ff.). Letzteres ist dabei ein Erkenntnisgewinn, der die anderen, praktischen Vorteile von Mappings ergänzt. Das theoretische Mapping von UWMETADATA nach MODS, so halten sie kurz fest, wurde durch die neue Architektur von Fedora 4 notwendig, das in dieser Version auf einen RDF-Standard wechselte. Die PHAIDRA-Arbeitsgruppen arbeiten daher daran, ein neues Datenmodell in RDF-Tripeln zu entwickeln, wodurch auch zahlreiche externe Vokabulare verwendbar werden. RDF-Graphen können außerdem problemlos erweitert werden und ermöglichen dadurch mehr Flexibilität, was das Metadatenschema an zukünftige Bedarfe anpassbar und interoperabel macht (vgl. Bellotto & Bettella 2019:182f.).

Fiala und Huggle (2019) beschreiben den Mapping-Prozess des Metadatenformats UWMETADATA, das in PHAIDRA, dem digitalen Repositorium der Universität Wien verwendet wird, nach MODS 3.6., das von der Library of Congress entwickelt wurde. Ausgangspunkt ist das Verstehen der Elemente der beiden Metadaten-Schemata und daran anschließend, als zweiter Schritt, das Mapping. Diesem Mapping nach MODS 3.6. sollte ein Mapping nach RDF folgen. Fiala und Huggle erwähnen die Ermöglichung des Datenaustausches in der Zukunft (vgl. Fiala & Huggle 2019:2) als Vorteil des Mappings.

Der Schwerpunkt dieser Arbeiten liegt auf der sehr genauen Beschreibung einzelner Mapping-Prozesse. Dabei zeigt sich an den individuellen Fällen, was durch Mappings

möglich gemacht werden kann: Sammlungen können von einem Einstiegspunkt aus durchsucht werden (vgl. Barroso u. a. 2015:43; Southwick 2015:2), dadurch können die digitalen Objekte einer größeren Nutzer*innen-Gruppe zugänglich gemacht werden (vgl. Southwick 2015:2). Die Arbeit mit Linked Open Data macht es darüber hinaus möglich, die eigenen Metadaten mit den Standards anderer Organisationen in Beziehung zu setzen (vgl. Southwick 2015:2). Außerdem können Mappings einen besseren Datenaustausch in der Zukunft ermöglichen (vgl. Fiala & Huggle 2019:2). Eine genauere Untersuchung der Umstände, die die Mappings notwendig machen und Vorteile, die sie bringen, wird jedoch nicht angestrebt und auch die erkennbaren Vorteile der Mappings werden nicht umfassend dargestellt. Nur Bellotto und Bettella (2019) halten in ihrer Fallstudie zum Mapping von UWMETADATA nach Dublin Core und MODS in PHAIDRA an der Universität Padua wichtige Vorteile, die eine Mapping bringen kann, in Kürze fest. Jedoch gehen sie dabei nicht im Detail auf die unterschiedlichen Punkte, die sie nennen ein. Diese Masterarbeit möchte die Arbeit von Bellotto und Bettella mit Bezug auf weitere in PHAIDRA verwendete Metadaten-Schemata tiefer analysieren, ergänzen und ausarbeiten.

- Ergebnisse des Projekts e-Infrastructures Austria im Bereich des Clusters I: Metadaten aus nicht-technischer und technischer Sicht

Das Projekt e-Infrastructures Austria setzte sich österreichweit zum Ziel, notwendige Infrastrukturen für digitale Repositorien aufzubauen und deren Ausbau und Entwicklung zu unterstützen. Dies sollte die Möglichkeit schaffen, digitale Objekte wie etwa elektronische Publikationen oder auch Multimedia-Dateien, die aus und für Lehre und Forschung entstanden, bereitzustellen und langfristig zu archivieren. Angeregt wurde es 2014 vom österreichischen Bundesministerium für Wissenschaft, Forschung und Wirtschaft und lief bis 2016 (vgl. Universität Wien 2023a). Der Work-Package-Cluster I setzte sich vertieft mit Metadaten auseinander. Die wichtigsten Ergebnisse in Form der Deliverables sollen nun vorgestellt und werden.

Für das Deliverable „Umgang mit Metadaten in Österreich“ untersuchten Blumesberger, Preza und Zartl den Umgang mit Metadaten in Repositorien und angedachte Entwicklungen in diesem Bereich. Dafür wurde ein schriftlicher Fragebogen erstellt und an

die Interview-Teilnehmer*innen der 25 Partnerorganisationen des Projekts e-Infrastructures Austria verschickt. Der Interviewleitfaden wurde dann während eines telefonischen Interviews durchgegangen. Der Vorteil dieser Vorgehensweise war, dass wenn etwas falsch oder nicht verstanden wurde, es gleich von beiden Seiten angesprochen und geklärt werden konnte. Das telefonische Gespräch wurde in Stichworten notiert, danach wurden Ergänzungen hinzugefügt. Die Fragebögen wurden archiviert, aber nicht an die Öffentlichkeit weitergegeben und die Verarbeitung fand anonym statt (vgl. Blumesberger, Preza, u. a. 2016:2).

Das Ergebnis der Befragung zeigte, der Umgang mit Metadaten war zum Zeitpunkt der Umfrage noch keine Selbstverständlichkeit oder Routine bei den Partnerorganisationen. Die Arbeit mit digitalen Repositorien war für viele der befragten Institutionen eine neue Aufgabe. Auch wenn die Arbeit mit traditionellen bibliothekarischen Metadaten vertraut war, war ihre Anwendung auf digitale Objekte noch nicht selbstverständlich. Für diese Anwendung braucht es jedoch nicht nur technische Kenntnisse, sondern auch das Erkennen der Relevanz von guten Metadaten in digitalen Repositorien für Forscher*innen und mögliche Nutzer*innen (vgl. Blumesberger, Preza, u. a. 2016:2).

Das Deliverable „Use of Metadata in Research Data Repositories: Use Case IST Austria DataRep" betrachtet die Verwendung von Metadaten für Forschungsdaten anhand des digitalen Repositoriums des Institute for Science and Technology Austria (IST), das der Speicherung von Forschungsdaten dient. Das Repositorium des IST verwendet Dublin Core als Metadaten-Schema ohne daran Anpassungen vorzunehmen (Krexhammer u. a. 2016:6ff.) und gliedert die Darstellung der Metadaten-Kategorien in drei Abschnitte: general, available files und additional information (vgl. Krexhammer u. a. 2016:8).

Das Deliverable „Metadaten und Forschungsdaten" untersuchte den Zusammenhang zwischen Metadaten und Forschungsdaten. Die Zunahme an digitalen Forschungsdaten macht deren Langzeitarchivierung notwendig, damit diese Forschungsdaten langfristig verfügbar sind und Forschungsergebnisse überprüft, Daten weiterverwendet, Langzeitstudien durchgeführt und interdisziplinäres Arbeiten möglich gemacht werden können. Auch wird auf Fördergeber verwiesen, die inzwischen Datenmanagementpläne fordern (vgl. Gstrein u. a. 2016:4). Die langfristige Speicherung von Forschungsdaten

kann dabei unterschiedliche Zeitspannen umfassen. Außerdem ist zu beachten, dass sich technische Umstände sowie Nutzungsumgebungen immer wieder ändern können. Daher ist es notwendig, eine adäquate und standardisierte Beschreibung der Objekte sicherzustellen, die die Vergabe von administrativen, technischen und deskriptiven Metadaten einschließt, um auch in Zukunft Emulation und Migration von, sowie Zugang zu digitalen Objekten zu ermöglichen. Um die Forschungsdaten strukturiert zu beschreiben, wenden verschiedene Disziplinen verschiedene Metadaten-Standards an. Da auch immer häufiger interdisziplinär gearbeitet wird, kann eine eindeutige Trennung zwischen den Fachbereichen und damit den relevanten Metadaten-Schemata nicht immer erfolgen. Dies wird auch durch die dargestellte Auswahl an Standards, die für die Sozialwissenschaften, Geisteswissenschaften, Altertums-wissenschaften und Geowissenschaften angeführt werden, deutlich (vgl. Gstrein u. a. 2016:6–9).

Das Deliverable „Metadata Quality“ widmet sich der Metadatenqualität, die ein entscheidender Faktor ist, um Daten und digitale Objekte leicht zu finden, sowie Zugang zu digitalen Objekten und Daten zu ermöglichen (vgl. Preza 2016:5). Zunächst wird in diesem Deliverable definiert, was Qualität in Bezug auf Metadaten bedeutet: Es können dafür die Vollständigkeit (engl. *completeness*), Genauigkeit (engl. *accuracy*), die Beschreibung der Entstehung (engl. *provenance*) der Metadaten betrachtet werden. Ein weiterer wichtiger Aspekt ist die Gestaltung der Metadaten entsprechen der Erwartungen der Community (engl. *conformance to expectations*), die sie verwendet. Außerdem sind logische Konsistenz und Kohärenz (engl. *logical consistency and coherence*), Aktualität (engl. *timeliness and currency*) der Metadaten, sowie ihre Zugänglichkeit (engl. *accessability*) wichtige Aspekte (vgl. Preza 2016:6f.). Es wird auch festgehalten, was ein Best-Practice-Ansatz für die Sicherstellung der Datenqualität sein könnte, der den Prozess vor (Archivierungsmöglichkeiten schaffen, Metadaten-Standards festlegen, Verantwortlichkeiten für die Qualitätssicherung zuweisen), während (Eingabehäufigkeit eines Datensatzes minimieren; einheitliche Terminologie verwenden, Granularität der Daten sicherstellen, Modifikationen am Datensatz festhalten) und nach (Suche nach Fehlern oder Unregelmäßigkeiten in den Daten) der Datensammlung umfasst. Evaluierung von Metadaten sollte durch ihre Analyse, die Suche nach Unregelmäßigkeiten und fehlenden Werten sowie die Validierung der verwendeten Schemata stattfinden (vgl. Preza 2016:8). Für die

Verbesserung der Metadaten-Qualität sollten diese gut dokumentiert sein, Standards und Best-Practice-Guidelines verwendet, lokale aber auch Community-Vokabulare veröffentlicht und gepflegt, sowie Application Profiles eingesetzt und Anleitungen verfasst werden (vgl. Preza 2016:10).

Das Deliverable „Metadata and Accessibility" untersucht, ob Metadaten die Barrierefreiheit von Objekten, die sie beschreiben, darstellen können. Dies soll dazu dienen Menschen mit Beeinträchtigungen, die Dokumente, die sie brauchen, auch zugänglich zu machen (vgl. Jeitler & Blumesberger 2016:4). Der Begriff Barrierefreiheit kann dabei unterschiedlichste Zugangsmöglichkeiten zu Objekten umfassen. Daher sind die Bedürfnisse, die durch „a person's ability to use, perceive or find information due to a specific impairment" (Jeitler & Blumesberger 2016:5) entstehen, sehr unterschiedlicher Natur und können schwer zu vergleichen sein (vgl. Jeitler & Blumesberger 2016:6). In diesem Deliverable wird festgehalten, dass ein Ausgangspunkt für barrierefreie Metadaten die vier Prinzipien der Web Content Accessibility Guidelines 2.0 (kurz WCAG 2.0) sein könnten, die jedoch für Metadaten angepasst werden sollten. Metadaten müssen dabei immer aus ihrem Kontext heraus betrachtet werden, d.h. auch in ihrer Darstellung im User Interface des Repositoriums. Darüber hinaus sollten Metadaten nicht nur barrierefrei zugänglich sein, sondern auch die Barrierefreiheit von Objekten beschreiben, damit Menschen beurteilen können, ob ein Dokument ihren Bedürfnissen gerecht wird (vgl. Jeitler & Blumesberger 2016:9). Diese Metadaten zur Barrierefreiheit können zusätzliche Funktionen, die Dokumente haben, und auch Gefahren, die von Dokumenten ausgehen beschreiben, etwa von blitzenden Lichtern bei Epilepsie. Darüber hinaus können diese Metadaten Schnittstellen für barrierefreie Technologien wie etwa Bildschirmlesegeräte benennen und angeben, auf welche Art auf die Dokumente zugegriffen werden kann, etwa über die Tastatur, die Maus, die Stimme (vgl. Jeitler & Blumesberger 2016:10).

Als zukünftiger Schritt ist ein gemeinsames Repositorium mit digitalen Objekten für Menschen mit bestimmten Behinderungen angedacht. Dafür soll untersucht werden, ob Accessibility Metadata des Projekts Schema.org/ Accessibility Metadata Project den Bedürfnissen der Menschen mit Behinderungen entsprechen kann. Daraufhin soll eine Entscheidung für ein bestimmtes Set an Metadaten getroffen werden und Infrastruktur

aufgebaut werden, die unterschiedlichen Versionen eines Objekts gerecht wird. Außerdem sollen Tools eingesetzt werden, die es Nutzer*innen mit Behinderung möglich machen, ihre Bedürfnisse zum Ausdruck zu bringen (vgl. Jeitler & Blumesberger 2016:11).

Das Deliverable „Jobprofil von MetadatenmanagerInnen“ hält die Aufgaben von Metadatenmanager*innen fest. Dazu gehören das Management von Metadaten, die Verwendung internationaler Datenstandards, die Fähigkeit Informationen zur Relevanz von guten Metadaten in digitaler Zeit bekanntzumachen, die Teilnahme an internationalen Kooperationen mit Expert*innen, das Beobachten von technischen Entwicklungen wie etwa von Linked Data und dem Semantic Web. Darüber hinaus Publikationstätigkeit, Erstellung von Guidelines und das Abhalten von Schulungen, die Auswahl von Metadaten-Schemata für Austausch von Daten und Langzeitarchivierung. Metadaten-Manager*innen sollen außerdem für alle Fragen, die das Thema Metadaten berühren, Ansprechperson sein, sie sollen einen Beitrag leisten zu besseren Suchergebnissen und zum Austausch von Metadaten zwischen verschiedenen Systemen. Auch die Mitarbeit in Kommissionen, Arbeitsgruppen und die Mitwirkung beim Forschungsdatenmanagement ist erwünscht. Es werden auch fachliche Fähigkeiten, Soft Skills und vorteilhafte Anforderungen für das Profil von Metadaten-Manager*innen aufgelistet: Zu den fachlichen Anforderungen gehören neben der Reifeprüfung auch Kenntnisse relevanter bibliothekarischer Regelwerke und Datenformate, Verständnis für technische Aspekte des Berufes, Genauigkeit, fortgeschrittene Englischkenntnisse und Projektmanagement-Erfahrung (vgl. Blumesberger, Gründhammer, u. a. 2016:5). Als Soft Skills werden eine starke kommunikative und soziale Kompetenz, Freude an Kontakten, Serviceorientiertheit, Teamfähigkeit, eine flexible Persönlichkeitsstruktur, die Bereitschaft sich immer wieder für neue Themen zu interessieren und sich weiterzubilden sowie Themen vor Publikum zu präsentieren und Reisebereitschaft genannt. Weitere vorteilhafte Aspekte für den Beruf der Metadaten-Manger*in sind der Abschluss der Universitätslehrgangs Library and Information Studies, oder eine andere relevante Ausbildung im Bibliothekswesen, juristische Kenntnisse, Kenntnisse und Fähigkeiten im Bereich XML und Linked Data, Ontologien und Klassifikationen sowie Erfahrung mit Datenmigration, -konversion und Datenmappings (vgl. Blumesberger, Gründhammer, u. a. 2016:6).

Das Deliverable „Classification Server“ beschreibt den Prozess der Implementierung eines Klassifikationsservers für ein digitales Repositorium (vgl. Kopácsi u. a. 2016:3) anhand des digitalen Repositoriums PHAIDRA der Universität Wien. Ein Klassifikationsserver dient dabei dazu, alle relevanten Klassifikationen, die standardisierten Begriffe von unterschiedlichen Arten von Klassifikationen oder kontrollierten Vokabularen zu versammeln und sie für die Nutzer*innen zugänglich und auswählbar zu machen (vgl. Kopácsi u. a. 2016:5). Das Deliverable beschreibt zunächst die allgemeinen und technischen Ansprüche, die ein Klassifikationsserver erfüllen sollte, um dann unterschiedliche Klassifikationsserver, (ThManager, TemaTres, SKOS[5] Shuttle, Poolparty, Protégé und SKOSMOS) zu analysieren (vgl. Kopácsi u. a. 2016:8–11). Die Entscheidung für SKOSMOS fällt, da SKOSMOS ein Open-Source-Tool ist, etwa Anfragen der Abfragesprache SPARQL annehmen kann und ein einfacher Zugang zum Klassifikationsserver über eine Programmierschnittstelle möglich ist. Anschließend wird der Prozess der Implementierung des Klassifikationsservers (vgl. Kopácsi u. a. 2016:11–18) und der Einsatz von möglichen kontrollierten Vokabularen festgehalten (vgl. Kopácsi u. a. 2016:18f.). Abschließend wird beschrieben, wie PHAIDRA mit dem Klassifikationsserver verbunden wird, um den Nutzer*innen bei der Eingabe von neuen digitalen Objekten oder der Suche danach, die kontrollierten Vokabulare zugänglich zu machen (vgl. Kopácsi u. a. 2016:19).

Zusammenfassend lässt sich für dieses Kapitel sagen, es gibt zwar Studien, die die Merkmale verschiedener Arten von KOS vergleichen, sowie es auch Untersuchungen dazu gibt, welche methodischen Ansätze für Mappings zwischen unterschiedlichen Arten von KOS verwendet werden können. Es gibt auch Berichte zu einzelnen Metadaten-Mappings, die konkret durchgeführt wurden, jedoch werden in keinem der drei Bereiche die Gründe betrachtet, die die Mappings notwendig machen.

Aus der Zusammenfassung der unterschiedlichen Ergebnisse des Work-Package-Cluster I „Metadaten aus technischer und nicht-technischer Sicht“ wird deutlich, für wie viele unterschiedliche Bereiche Metadaten Relevanz haben. In diesem Arbeitspaket

[5] SKOS steht für: Simple Knowledge Organization System

wurde der Umgang mit Metadaten in unterschiedlichen Repositorien betrachtet, der Zusammenhang zwischen Meta- und Forschungsdaten beleuchtet und Parameter, die zur Qualität von Metadaten beitragen können, versammelt. Außerdem wurde untersucht, wie Metadaten barrierefrei gestaltet werden können und barrierefrei Dokumente beschreiben können. Auch wurde ein Job-Profil für Metadaten-Manager*innen entwickelt und die Implementierung eines Klassifikationsservers beschrieben. Dabei wurde jedoch nicht darauf eingegangen, was geschieht, wenn Metadaten bereits vorhanden sind und in ein anderes Format gemappt werden müssen.

2. Wissensorganisation und Library and Information Studies

Dieses Kapitel befasst sich mit dem Zusammenhang zwischen Wissensorganisation und Library and Information Studies. Zunächst wird der Bereich der Wissensorganisation vorgestellt und definiert, um herauszuarbeiten, dass es sich bei der Arbeit im Bereich der Metadaten um einen Teilbereich der Wissensorganisation handelt. Anschließend wird der Begriff *Metadaten* im Kontext seiner jahrtausendealten Geschichte analysiert. Es werden Definitionen dieses Begriffs in der Gegenwart untersucht und die Verwendung des Begriffs *Metadaten* im Bereich Library and Information Studies betrachtet. Abschließend wird PHAIDRA, das digitale Repositorium der Universität Wien als ein konkretes Beispiel eines digitalen Repositoriums – und der damit zusammenhängenden Arbeit mit Metadaten – vorgestellt, das im Zentrum der Analyse dieser Arbeit steht.

2.1. Der Zusammenhang zwischen den Bereichen Wissensorganisation und Library and Information Studies

Wissensorganisation (engl. *knowledge organization*) ist ein Forschungsgebiet und eine Disziplin, die unterrichtet und praktisch angewandt wird. Sie ist eng mit dem Fachbereich Library and Information Science verbunden (Hjørland & Gnoli 2016). Gnoli (2020:6) bietet eine recht weite Definition des Begriffs *Wissensorganisation*: "knowledge organization means the representation of real objects (knowledge) arranged into useful structures (organization)". Gnoli macht es hier möglich, die Organisation der unterschiedlichsten Objekte in den unterschiedlichsten Arten von Strukturen als Wissensorganisation zu betrachten, dazu können etwa auch Webseiteninhalte eines Fernsehsenders zu einer TV-Serie gehören (vgl. Gnoli 2020:6).

Die Vielfalt, die im Begriff Wissensorganisation liegt, beschreibt auch Hjørland (2008:86). Er definiert den Begriff Wissensorganisation in einem engeren und in einem weiteren Sinne: Die engere Definition wird von Hjørland als praktische Aufgabe der Wissensorganisation gesehen: „activities such as document description, indexing and classification performed in libraries, bibliographical databases, archives and other kinds of ‚memory institutions' by librarians, archivists, information specialists, subject specialists,

as well as by computer algorithms and laymen“ (Hjørland 2008:86). Damit wird deutlich, dass Wissensorganisation viele Aktivitäten in Gedächtnisinstitutionen wie etwa Bibliotheken und Archiven umfasst. Davon ausgehend beschäftigt sich die Wissensorganisation als Forschungsdisziplin mit der Art wie diese Tätigkeiten gestaltet werden und den für sie relevanten Qualitätskriterien. Außerdem betrachtet die Wissensorganisation, wie und welche Wissensorganisationssysteme verwendet werden, um Dokumente, ihre Darstellungen und Begriffe zu strukturieren (vgl. Hjørland 2008:86).

Die zweite, weitere Definition der Wissensorganisation umfasst diese als Bereich, der sich mit der „social division of mental labor“ (Hjørland 2008:86) befasst: also der Art, wie intellektuelle Arbeit in gesellschaftlichen Gefügen verteilt wird. Hjørland führt hier als Beispiel die Organisation von Forschungs- und tertiären Bildungsinstitutionen an, sowie die Strukturierung von Fachbereichen und auch die Prozesse der Produktion und Verbreitung von Wissen, wie sie auch von Gnolis Definition umfasst wären. Hjørlands engere Definition der Wissensorganisation bezieht sich daher stärker auf die kognitive Strukturierung und Darstellung von Wissen, während die weitere Definition soziale Aspekte der Organisation von Wissen miteinbezieht (vgl. Hjørland 2008:86).

Zum engeren Begriff der Wissensorganisation Hjørlands können unterschiedliche Arten von Wissensorganisationssystemen gehören: so etwa die Begriffe kontrollierter Vokabulare, Listen von Begriffen oder hierarchische Begriffsstrukturen, sowie Taxonomien, Schlagwortkataloge, Thesauri, Klassifikationen und auch die zeitgenössischen Ontologien (vgl. Gnoli 2020:74–85). Auch Metadaten/Metadatenformate sind Wissensorganisationssysteme. Gnoli bemerkt, dass die Vergabe von Metadaten „just another form of traditional indexing“ ist (Gnoli 2020:85). Die Vergabe von Metadaten ist daher eine Aufgabe der Wissensorganisation. Hier hat eine altbekannte bibliothekarische Tätigkeit einen neuen Namen erhalten. Für diese Arbeit soll die engere Definition des Begriffs Wissensorganisation herangezogen werden, da diese sich auf die kognitive Strukturierung und Darstellung von einer bestimmten Art von KOS, den Metadaten, beziehen lässt.

2.2. Metadaten

Metadaten werden häufig kurz als „data about data“ (Gnoli 2020:85) bezeichnet. Strukturierten Metadaten begegnet man im Alltag sehr häufig. Sie dienen dazu Daten zu beschreiben und treten immer gekoppelt an diese auf. Zeng und Qin (2022:4) führen als Beispiel dafür die Nährwertangaben und Inhaltsstoffe auf Nahrungsmitteln und Getränken an. Die Beschreibung der Inhaltsstoffe und Nährwerte durch Metadaten „uses property-value paired statements to clearly describe the characteristics of a specific thing. A property specifies the meaning of a data value“ (Zeng & Qin 2022:4). Angewandt auf die Nährwertangaben bedeutet dies, die Angabe des Merkmals (engl. *property*) *Kalorien* auf der Müslipackung verleiht dem Wert (engl. *value*), etwa *230*, erst eine Bedeutung, die der Mensch verstehen kann. Im Alltag gibt es zahlreiche Gebiete, für die Metadaten relevant sind. So sind etwa Medikamente mit Metadaten-Daten-Paaren zu Inhaltsstoffen und zu Nebenwirkungen versehen. Darüber hinaus begegnet man ihnen in Patientenakten, Steuerformularen, Zeugnissen und Online-Profilen. Diese Angaben können natürlich komplexer sein als die auf Nahrungsmitteln (vgl. Zeng & Qin 2022:4f.).

Die Wissensorganisation als bibliothekarische Aktivität gibt es schon sehr lange. Sie zeigt sich in frühen menschlichen Aktivitäten, die den Wunsch belegen, Wissen zu strukturieren. Gnoli (2020:10) verweist hier auf Goody (1977) und führt als ein sehr frühes Beispiel für Tätigkeiten, die der Wissensorganisation zugeordnet werden können, die ältesten erhaltenen schriftlichen Dokumente der Sumerer aus der Zeit um 3000 v. Chr. an: Auf Lehmtafeln und gewachsten Holzbrettern wurden damals erste Listen erstellt und damit Wissen strukturiert. Listen sind dabei eine grundlegende Art von Wissensorganisationssystemen. Etwas später wurden auch schon Metadaten vergeben, als die Bibliothek des Aššurbanipal in Ninive um etwa 650 v. Chr. entstand, die eine Vielzahl wichtiger Texte, wie etwa eine Abschrift des Gilgamesh-Epos auf Lehmtafeln enthielt. Die Tafeln waren thematisch aufgestellt, in Gefäßen, die auf Regalen in Lagerräumen standen. Sie wurden teilweise auch katalogisiert und auf diese Weise mit Metadaten versehen. Denn die Wände und Türen einiger Räume enthielten eingeritzte Angaben dazu, welche Titel wo enthalten waren, wie Gnoli (2020:10) mit Verweis auf Hyman (1982:7) berichtet. Mit

dieser Art der Katalogisierung wurden schon sehr früh Daten (die Einkerbungen an den Türen und Wänden) über Daten (die Tafeln) erfasst und damit Metadaten vergeben.[6]

Die erwähnte sehr kurze Definition Metadaten seien „data about data" oder auch in abgewandelter Form, sie seien „information about information" (Zeng & Qin 2022:12) kann sehr viele Bereiche und Darstellungsarten umfassen. Zeng und Qin verweisen hier auf die FGDC (2000), Hodge (2001) und Ri (2004), die diese beiden Definitionen ebenfalls verwenden.

In Bibliotheken und Archiven und Museen (engl. *libraries, archives and museums* und daher kurz LAM-Institutionen) können Metadaten Literatur oder Museumsobjekte beschreiben. Genauso können sie dazu dienen, Materialien in Archiven oder wissenschaftliche Zeitschriftenartikel über ihre Indexierung in Datenbanken zu finden (vgl. Zeng & Qin 2022:12). Zeng und Qin formulieren eine spezifischere Definition, um die Arbeit mit Metadaten in LAM-Institutionen zu erfassen. Für Zeng und Qin (2022:14) sind Metadaten „Structured, encoded data that describe characteristics of information-bearing entities to aid in the identification, discovery, assessment, management, and preservation of the described entities" Die Anforderungen an Metadaten in LAM-Institutionen sind somit ausdifferenziert: Metadaten müssen hier auf strukturierte und formalisierte Art und Weise dargestellt sein. Sie sollen dabei die Merkmale derjenigen Objekte festhalten, die Informationsträger sind. Indem sie diese Aufgabe übernehmen, ermöglichen sie es auch,

[6] Wissensorganisation fand sich in vielen Kulturen schon sehr früh. Um einen kleinen Einblick zu bieten, sollen hier auch Ansätze anderer früher Kulturen genannt werden: Im alten Ägypten kannte man ebenfalls strukturierte Listen, die sie um 2500 BCE zunächst auf Lehmtafeln und später dann auf Papyrusrollen festhielten. Später wurden die verwendeten Vokabulare besser entwickelt, wie etwa im *Ramesseum*, einem ägyptischen Lexikon, in dem Worte formalisiert dargestellt und mit klassifizierenden Überschriften versehen sind, erläutert Gnoli (2020:11) und verweist auf Gardiner (1947).

Auch in Indien nutzte man früh Prinzipien der Wissensorganisation, um die religiösen Texte der Veden zu strukturieren, die zwischen 1500-800 BCE in Sanskrit verfasst wurden (vgl. Gnoli 2020:11f.). Gnoli verweist auf die beiden bedeutenden indischen Bibliothekare Bhattacharyya and Ranganathan (1967:121–2). Jedoch hatte in den Veden das Strukturierungsprinzip eine moralische Basis, die die Entwicklung der Menschen zum Ziel hatte (vgl. Gnoli 2020:11f.). Es wurde weniger zur allgemeinen, eher nüchternen Strukturierung von Wissen verwendet.

Die chinesische Philosophie unterscheidet traditionellerweise zwischen den beiden gegensätzlichen Prinzipien (oder auch Kategorien) Yin und Yang: wobei Yin oft durch eine gestrichelte Linie und Yang oft durch eine durchgehende Linie dargestellt wird. Im Buch der Wandlungen werden diese beiden Darstellungsarten kombiniert, weiter unterteilt und übereinander dargestellt, um alle Phänomene der Welt zu erfassen. Laut Gnoli wird hier durch die Verbindung von zwei grundlegenden Unterscheidungselementen eine Methode eingesetzt, die auch in der Wissensorganisation häufig Anwendung findet (vgl. Gnoli 2020:12).

diese Objekte in einer Menge weiterer Objekte zu identifizieren, zu finden, sie zu bewerten und zu organisieren, um mit ihnen arbeiten zu können. Dies gilt nicht nur für physische, sondern auch für digitale Objekte in digitalen Repositorien. Letztere ähneln in einem grundlegenden Aspekt ihres Aufbaus dem Aufbau der Bibliothek des Aššurbanipal in Ninive: Es gibt einen – wenn auch digitalen – Raum, in dem – nun digitale – Objekte gelagert werden, die mit Metadaten versehen sind, um sie finden zu können. Die digitalen Objekte werden zwar mit komplexeren und stärker strukturierten und formalisierten Metadaten versehen, die Grundfunktion bleibt jedoch die gleiche.

2.3. PHAIDRA – das digitale Repositorium der Universität Wien

Einer dieser digitalen Räume, in dem digitale Objekte gelagert und mit Metadaten versehen werden, ist PHAIDRA (Akronym für Permanent Hosting, Archiving and Indexing of Digital Resources and Assets), das digitale Repositorium der Universität Wien. Es steht mit seinen Metadatenformaten im Zentrum dieser Arbeit und soll daher nun vorgestellt werden. PHAIDRA dient der langfristigen Sicherung digitaler Objekte und Sammlungen (vgl. Universität Wien 2023b). Dazu gehören digitale Objekte der Forschung, der Lehre und auch der Verwaltung. PHAIDRA ermöglicht diesen Bereichen die „langfristige […] Verwaltung von Digitalisaten, eine flexible Nutzung der Ressourcen, sowie das Auffinden und Abrufen von vorbereiteten digitalen Objekten“ (Universitätsbibliothek Universität Wien 2023). Dazu gehören „Forschungsergebnisse, Projektberichte, Publikationen, Lehrmaterialien, Digitalisate, Bilder und vieles mehr“ (Universitätsbibliothek Universität Wien 2023).

Für die Speicherung dieser digitalen Objekte sichert PHAIDRA „Integrität, Authentizität, Erhaltung der Interpretierbarkeit, Vertraulichkeit und Verfügbarkeit“ zu (Universitätsbibliothek Universität Wien 2023) . Die Daten werden langfristig gesichert, sie werden systematisch festgehalten, mit Metadaten annotiert und sind ohne zeitliche Begrenzung global verfügbar. Außerdem erhalten sie einen permanenten Link, der sie dauerhaft zugänglich macht (vgl. Universitätsbibliothek Universität Wien 2023).

Unter den Vorteilen, die PHAIDRA bietet, werden die Metadaten-Schemata genannt, die auf Dublin Core und LOM (kurz für *Learning Object Metadata*) basieren und

aus acht verpflichtenden Feldern bestehen. Diese können zusätzlich flexibel mit weiteren Metadatenkategorien ergänzt werden (vgl. Universität Wien 2023b). Die Daten, also die digitalen Objekte, die in PHAIDRA gespeichert werden, sollen den FAIR-Prinzipien entsprechen und damit „findable, accessible, inter-operable und re-usable“ (Universität Wien 2023b) sein.

In diesem Kapitel wurde betrachtet, wie die beiden Disziplinen der Wissensorganisation und der Library and Information Studies miteinander verbunden sind. Der Bereich der Wissensorganisation kann dabei die Arbeit mit bibliothekarischen Metadaten umfassen. Es wurde gezeigt, dass sowohl die Wissensorganisation als auch die Vergabe von Metadaten eine jahrtausendealte Tradition sind, die sich erhalten hat und im digitalen Zeitalter ein zentraler Bestandteil der Library and Information Studies sind. Darüber hinaus wurden aktuelle Definitionen und Verwendungen des Begriffs *Metadaten* im Bereich Library and Information Studies untersucht. Abschließend wurde PHAIDRA, das digitale Repositorium der Universität Wien inklusive seiner Zielsetzung vorgestellt, da es ein konkreter Anwendungsfall der Arbeit mit Metadaten ist. Die Arbeit mit Metadaten in PHAIDRA ist Gegenstand der Analyse dieser Arbeit.

3. Prinzipien für gute Metadaten: die Prinzipien des „Framework of Guidance for Building Good Digital Collections“ und die FAIR-Prinzipien

In diesem Kapitel werden als Ausgangspunkt für die Analyse dieser Arbeit das „Framework of Guidance for Building Good Digital Collections“ der *National Information Standards Organization* (kurz NISO) und seine sechs Prinzipien für gute Metadaten vorgestellt. Anschließend werden die aktuell viel beachteten FAIR-Prinzipien für gute (Meta-)Daten präsentiert und mit den NISO-Prinzipien verglichen.

3.1. Das “Framework of Guidance for Building Good Digital Collections” der National Information Standards Organization und seine Metadaten-Prinzipien

Die *National Information Standards Organization* ist eine US-amerikanische Non-Profit-Organisation, die es als ihre Aufgaben definiert, Standards und Empfehlungen zu erarbeiten, zu identifizieren, zu pflegen, weiterzuentwickeln und zu veröffentlichen. Diese Standards und Empfehlungen sollen es ermöglichen, in der sich immerzu verändernden digitalen Welt Informationen besser zu managen. Die Standards, die dabei entwickelt werden, beziehen die Arbeitsweisen sowohl traditioneller als auch neuer Technologien mit ein, können auf den gesamten Lebenszyklus von Informationen angewandt werden und umfassen Phasen „from creation through documentation, use, repurposing, storage, metadata, and preservation“ (NISO 2023).

Das „Framework of Guidance for Building Good Digital Collections“ (NISO 2007) befasst sich mit digitalen Sammlungen generell, die von Gedächtnisinstitutionen unterschiedlichster Art, von Museen über Bibliotheken und Archive bis hin zu Schulen, erstellt werden, und soll Orientierungspunkte für den Umgang mit ihnen bieten (vgl. NISO 2007, 2f.). Außerdem soll es auch der Unterstützung von Fördergebern dienen, die sich für die Entwicklung guter digitaler Sammlungen einsetzen wollen (vgl. NISO 2007:1). Dabei gilt generell:

> “There are no absolute rules for creating good digital collections. Every digital collection-building initiative is unique, with its own users, goals, and needs. Initiatives dealing with legacy collections, for example, have different constraints than projects embarking on new digitization efforts, which in turn have different constraints than projects building collections of born-digital materials. Museums, libraries, archives, and schools have different constituencies, priorities, institutional cultures, funding mechanisms, and governance structures. The key to a successful project is not to strictly and unquestioningly follow any particular path, but to plan strategically and make wise choices from an array of tools and processes to support the unique goals and needs of each collection.” (NISO 2007:2f.)

Die Vorschläge der NISO sind daher nicht präskriptiv zu verstehen, sondern eher als Leitlinien, die es ermöglichen sollen, den Bedarfen der jeweiligen Institution gerecht zu werden.

Für die Arbeit mit digitalen Sammlungen entwickelte die NISO-Gütekriterien für vier Bereiche. Diese sind:

„• Collections (organized groups of objects)
• Objects (digital materials)
• Metadata (information about objects and collections)
• Initiatives (programs or projects to create and manage collections)”
(NISO 2007:2 H.i.O.)

Der Begriff der Güte umfasst dabei unterschiedliche Aspekte, die, neben fachlichen und technischen Entwicklungen, auch soziale Entwicklungen miteinbeziehen. Digitale Sammlungen sollen sich daher etwa auch neuen Zielgruppen öffnen, wobei Kosten-Nutzen-Rechnungen aufgestellt, Nachhaltigkeit sichergestellt und digitale Sammlungen als vertrauenswürdigen Quellen zu identifizieren sein sollten. Die digitalen Sammlungen, die digitalen Objekte in ihnen und auch ihre Metadaten können nicht mehr nur in Bezug auf die Projekte, aus denen heraus sie entstanden sind, betrachtet werden. Sie müssen auch Objekte sein, die andere Nutzer*innen auf unterschiedlichste Arten wiederverwenden und weiterverarbeiten können (vgl. NISO 2007:1). Güte verlangt daher auch nach “interoperability, reusability, persistence, verification, documentation, and support for intellectual property rights” (NISO 2007:1). Für Metadaten im Speziellen gilt daher, sie sollten interoperabel sein, wiederverwendet werden können, langlebig sein, ihre Richtigkeit sollte sichergestellt werden, eine Dokumentation zu den verwendeten Metadaten verfügbar sein und sie sollten geistige Eigentumsrechte abbilden können (vgl. NISO 2007:1).

Die Gütekriterien der NISO beziehen sich auf digitale Objekte und Sammlungen, Metadaten und Projekte generell und nicht ausschließlich auf digitale Objekte, Sammlungen, Metadaten und Projekte, die Resultat eines Forschungsprozesses sind. Damit können sie prinzipiell auch Aspekte digitaler Objekte der Verwaltung erfassen, wie sie an der Universität Wien etwa wie durch das Universitätsarchiv in PHAIDRA abgelegt werden.[7]

Die Prinzipien der NISO für gute Metadaten lauten:

> „**Metadata Principle 1**: Good metadata conforms to community standards in a way that is appropriate to the materials in the collection, users of the collection, and current and potential future uses of the collection.
>
> **Metadata Principle 2**: Good metadata supports interoperability.
>
> **Metadata Principle 3**: Good metadata uses authority control and content standards to describe objects and collocate related objects.
>
> **Metadata Principle 4**: Good metadata includes a clear statement of the conditions and terms of use for the digital object.
>
> **Metadata Principle 5**: Good metadata supports the long-term curation and preservation of objects in collections.
>
> **Metadata Principle 6**: Good metadata records are objects themselves and therefore should have the qualities of good objects, including authority, authenticity, archivability, persistence, and unique identification." (NISO 2023:61f., H.i.O.)

Diese sechs Prinzipien werden in dieser Arbeit als Ausgangspunkt verwendet, anhand derer untersucht werden soll, wo die Gründe liegen könnten, die ein Mapping von einem Format in ein anderes notwendig machen können, und die auch durch weitere Aspekte ergänzt werden können.

Nachdem die NISO und ihre Prinzipien für gute Metadaten vorgestellt wurden, sollen im Folgenden die FAIR-Prinzipien betrachtet werden, um sie daraufhin mit den NISO-Prinzipien in Beziehung zu setzen und Gemeinsamkeiten und Unterschiede herauszuarbeiten.

[7] In Phaidra finden sich so Dokumente des Archivs der Universität Wien als digitale Objekte. Dazu gehören etwa Inskriptionsformulare von Frauen der Medizinische Fakultät von 1902 bis 1921 (Archiv der Universität Wien 2019) oder auch Bildersammlungen zu unterschiedlichen Anlässen (Maisel 2013).

3.2. Die FAIR-Prinzipien

Die FAIR-Prinzipien wurden erstmals in dem Artikel *FAIR Guiding Principles for scientific data management and stewardship* (Wilkinson u. a. 2016) veröffentlicht. Die Autoren gaben darin Empfehlungen und Orientierungspunkte, die es ermöglichen, digitale Objekte „Findable, Accessible, Interoperable and Reusable“ (Wilkinson u. a. 2016:3) zu machen. Diese Prinzipien entstanden aus dem Wunsch und der drängenden Notwendigkeit “to improve the infrastructure supporting the reuse of scholarly data.” (Wilkinson u. a. 2016:1). Damit konzentrieren sich die FAIR-Prinzipien auf vor allem wissenschaftliche Daten und Metadaten, die von und für Lehre, Wissenschaft und Forschung entstehen und archiviert werden sollen.

Ziel der FAIR-Prinzipien ist es, “discovery through good data management” (Wilkinson u. a. 2016:1) zu unterstützen und damit Innovationen sowie die weitere Verwendung und Integration von Daten und Wissen zu fördern, um die Investitionen, die in Forschung fließen, bestmöglich zu nutzen. Den bestmöglichen Nutzen aus Investitionen in die Forschung zu holen, motiviert(e) auch Fördergeber wissenschaftlicher Projekte, staatliche Stellen und Verlage, Datenmanagementpläne einzufordern, die die FAIR-Prinzipien berücksichtigen. Dazu gehört neben der Sammlung, Annotation und Archivierung von Daten auch die langfristige Pflege digitaler Objekte für die langfristige wissenschaftliche Wieder- und Weiterverwendung. Dies soll neue Entdeckungen fördern, indem Daten zugänglich gemacht werden, weiterverwendet und/oder mit weiteren (neuen) Daten für Forschungsvorhaben verwendet und/oder kombiniert werden können (vgl. Wilkinson u. a. 2016:1).

Die FAIR-Prinzipien beleuchten die vier in den Buchstaben ihrer Benennung versammelten Aspekte der Findability, Accessibility, Interoperability und des Reuse von Daten und Metadaten von digitalen Objekten für Forschung, Wissenschaft und Lehre (vgl. GO FAIR 2023a). Die vier grundlegenden Prinzipien der Findability, Accessibility, Interoperability und des Reuse sind in weitere Unterkategorien unterteilt. Die FAIR-Prinzipien finden sich inklusive ihrer Unterpunkte auf der folgenden Seite und lauten:

“Findability
F1: (Meta) data are assigned globally unique and persistent identifiers
F2: Data are described with rich metadata
F3: Metadata clearly and explicitly include the identifier of the data they describe
F4: (Meta)data are registered or indexed in a searchable resource

Accessibility
A1: (Meta)data are retrievable by their identifier using a standardised communication protocol
A1.1: The protocol is open, free and universally implementable
A1.2: The protocol allows for an authentication and authorisation procedure where necessary
A2: Metadata should be accessible even when the data is no longer available

Interoperability
I1: (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
I2: (Meta)data use vocabularies that follow the FAIR principles
I3: (Meta)data include qualified references to other (meta)data

Reusability
R1: (Meta)data are richly described with a plurality of accurate and relevant attributes
R1.1: (Meta)data are released with a clear and accessible data usage license
R1.2: (Meta)data are associated with detailed provenance
R1.3: (Meta)data meet domain-relevant community standards” (GO FAIR 2023a:H. durch Autorin)

Nachdem die Entstehung und Anwendungsbereiche der FAIR-Prinzipien sowie die FAIR-Prinzipien selbst vorgestellt wurden, sollen sie nun zu den NISO-Prinzipien für gute Metadaten in Beziehung gesetzt werden.

3.3. Die NISO-Prinzipien für gute Metadaten und die FAIR-Prinzipien im Vergleich

Es ist deutlich geworden, dass die NISO-Prinzipien für gute Metadaten und die FAIR-Prinzipien auf unterschiedliche Arten von Metadaten angewandt werden könnten: Die NISO-Prinzipien beziehen die Metadaten von verschiedensten Arten digitaler Sammlungen, und deren Metadaten mit ein, so auch Metadaten von administrativen digitalen Objekten. Die FAIR-Prinzipien dagegen sind aus dem Wunsch heraus entstanden, wissenschaftliche Daten und Metadaten aus Lehre, Forschung und Wissenschaft zu verbessern. Trotz des Unterschieds in der Zielsetzung findet sich eine gemeinsame Schnittmenge, wenn die beiden Gruppen von Prinzipien verglichen werden, wie nun gezeigt werden soll. Dafür werden die einzelnen Prinzipien der NISO vorgestellt und mit den relevanten

FAIR-Prinzipien verknüpft. Die Grundlage für den Vergleich bildet der gesamte Text der NISO (2007: 58-58) zu den Metadaten-Prinzipen und auch die ausführliche Beschreibung der FAIR-Prinzipien (2023a) auf ihrer Webseite. Wo es keine Parallelen zwischen den NISO-Prinzipien und den daraus abgeleiteten Kategorien zu den FAIR-Prinzipien gibt, wir dies auch festgehalten.

- Das erste Prinzip der NISO für gute Metadaten und die FAIR-Prinzipien

Das erste Prinzip der NISO für gute Metadaten regt an, Community Standards zu verwenden (vgl. NISO 2007:63–75). Dies findet sich im Wesentlichen auch in den FAIR-Prinzipien unter dem Punkt Reusability: „R1.3: (Meta)data meet domain-relevant community standards“ (GO FAIR 2023e). Damit entspricht das Metadaten-Prinzip 1 der NISO im Wesentlichen einem FAIR-Prinzip, wobei das FAIR Prinzip R1.3 noch betont, die ausgewählten Community Standards sollten für den jeweiligen Fachbereich relevant sein. Dies könnte jedoch auch in dem Metadaten-Prinzip der NISO impliziert sein, da eine Community sich meist auch innerhalb eines bestimmten Fachbereichs konstituiert.

- Das zweite Prinzip der NISO für gute Metadaten und die FAIR-Prinzipien

In ihrem zweiten Metadaten-Prinzip schlägt die NISO vor, Interoperabilität zu fördern (vgl. NISO 2007:76). Dieses Prinzip enthält vier Aspekte, die unterschiedliche Zugänge für die Förderung von Interoperabilität von Metadaten beleuchten. Der erste Aspekt ist die Vergabe von detaillierten Metadaten für digitale Objekte als Grundlage für Interoperabilität (vgl. NISO 2007:76). Die NISO nennt als Beispiel für die detaillierte Vergabe von Metadaten zwar die detaillierte Beschreibung von Abbildungen, jedoch können zu diesem Prinzip auch weitere Datenkategorien gehören, wie sie in Prinzip R1 der FAIR-Prinzipien enthalten sind, denn dort heißt es: „(Meta)data are richly described with a plurality of accurate and relevant attributes“ (GO FAIR 2023d). Ziel sowohl des NISO-Ansatzes als auch des FAIR-Prinzips, ist es, die Metadaten wiederverwendbar zu machen.

Ein weiterer Punkt, der in diesem zweiten Prinzip der NISO für gute Metadaten enthalten ist und sich in den FAIR-Prinzipien weiterentwickelt findet, ist die Anregung

Interoperabilität durch die Speicherung von Metadaten-Schemata in Repositorien zu ermöglichen (vgl. NISO 2007:77). Dies findet sich auf andere Weise in den FAIR-Prinzipien: Das Prinzip Interoperabilität regt unter dem Punkt I.1. an, Meta(daten) sollen „a formal, accessible, shared, and broadly applicable language for knowledge representation“ (GO FAIR 2023a) verwenden. Die Forderung diese Sprachen sollten „accesssible, shared“ (GO FAIR 2023a) also zugänglich sein, impliziert, sie könnten in einem Repositorium oder einem anderen geeigneten Ort abgelegt sein. Damit findet sich dieser Aspekt der NISO auch in den FAIR-Prinzipien.

Es finden sich jedoch nicht alle Aspekte des Prinzips der NISO, das Interoperabilität unterstützen möchte, in den FAIR-Prinzipien. Dazu gehören der Wunsch nach der Zusammenarbeit von Organisationen und Aggregatoren (vgl. NISO 2023:76), sowie die Erarbeitung von Mappings (vgl. NISO 2007:77). Auch die Ermöglichung von Metasuchen (vgl. NISO 2007:78) und die Verwendung von Harvesting-Protokollen (vgl. NISO 2007:78) werden in den FAIR-Prinzipien nicht erwähnt.

- Das dritte Prinzip der NISO für gute Metadaten und die FAIR-Prinzipien

Das Metadaten-Prinzip 3 der NISO, das zum Einsatz von standardisierten und kontrollierten Begriffen rät (vgl. NISO 2023:79), findet sich im FAIR-Prinzip I2., das anregt Vokabulare zu verwenden werden, die den FAIR-Prinzipien entsprechen (vgl. GO FAIR 2023a). Darüber hinaus regt die NISO in diesem dritten Prinzip auch die Dokumentation der verwendeten Begriffe an (vgl. NISO 2007:79). Dieser Aspekt ist implizit im FAIR-Prinzip I1 enthalten, das die Verwendung von standardisierten, weit verbreiteten und leicht zugänglichen Sprachen für Wissensrepräsentation und damit der Darstellung der (Meta-)Daten rät (vgl. GO FAIR 2023c). Die darin verwendeten Begriffe müssen durch ihre Formalisierung, die damit einhergehende Standardisierung und die leichte Zugänglichkeit und breite Verwendung auch gut dokumentiert sein.

- Das vierte Prinzip der NISO für gute Metadaten und die FAIR-Prinzipien

Das vierte Prinzip für gute Metadaten der NISO fordert die Festlegung der rechtlichen Nutzungsbedingungen von digitalen Objekten (vgl. NISO 2007:81). Eine Parallele dazu finden sich in dem FAIR-Prinzip R1 und seinen Unterpunkten, das in der Lage ist, dieselben Aspekte wie die NISO, dazu gehören die Beschreibung der Rechte, das Management und die Kommunikation von Rechten und die Verwaltung von rechtlich begründeten Zugangsbedingungen abzudecken. Das FAIR-Prinzip fordert dies darüber hinaus in einer Form, die sowohl von Menschen als auch Maschinen gelesen und verarbeitet werden kann (vgl. GO FAIR 2023c). Dies wird von der NISO zwar auch für das Management und die Kommunikation von Rechten im Bereich der Lizenzierungsbedingungen angeregt, aber nicht im selben Ausmaß, wie von den FAIR-Prinzipien gefordert (vgl. NISO 2007:81). Grund dafür ist vermutlich der Zeitpunkt der Entstehung einerseits der NISO-Prinzipien und der FAIR-Prinzipien. Wobei letztere knapp zehn Jahre nach der Entwicklung der NISO-Prinzipien publiziert wurden.

- Das fünfte Prinzip der NISO für gute Metadaten und die FAIR-Prinzipien

Im fünften Prinzip für gute Metadaten fordert die NISO, diese sollten zum langfristigen Erhalt und zur Pflege der digitalen Objekte beitragen. Dafür sollen technische Metadaten, Erhaltungsmetadaten (auch Retentionsmetadaten genannt), Recordkeeping-Metadaten und die strukturellen Metadaten eingesetzt werden (vgl. NISO 2007:83). Dieses Prinzip der NISO zeigt sich leicht modifiziert in dem FAIR-Prinzip *Reusable*, das im Prinzip R1 rät, (Meta)-Daten mit einer Vielzahl relevanter Merkmale zu beschreiben (GO FAIR 2023c), wozu auch die unterschiedlichen Metadaten-Arten, die von den Metadaten-Prinzipien der NISO genannt werden, gehören können.

- Das sechste Prinzip der NISO für gute Metadaten und die FAIR-Prinzipien

Das sechste Prinzip der NISO für gute Metadaten schlägt vor, Metadaten wie digitale Objekte zu behandeln und sie dieselben Kriterien und Prinzipien erfüllen zu lassen, wie sie im Kapitel zu digitalen Objekten der NISO beschrieben werden (vgl. NISO 2007:85). Dort wiederholen sich leicht modifiziert einige Aspekte, die auch in den Prinzipien der

NISO für gute Metadaten vorkommen, wie der Einsatz von Formaten, die eine langfristige Nutzung und Zugänglichkeit der Metadaten fördern (vgl. NISO 2007:26–57). Dieser Aspekt zeigen sich auch in dem bereits erwähnten FAIR-Prinzip I1, das den Einsatz von zugänglichen, weit verbreiteten, formalisierten und auf viele Bereich anwendbaren Sprachen der Wissensrepräsentation anregt (vgl. GO FAIR 2023a). Ein wichtiger, bisher nicht beachteter Punkt, ist der Einsatz von persistenten, eindeutigen Identifikatoren, der in den FAIR-Prinzipien unter *Findable* vorkommt. Dort heißt es in Prinzip "F1:(Meta)data are assigned globally unique and persistent identifiers" (GO FAIR 2023b).

Die FAIR-Prinzipien sind damit im Wesentlichen in den NISO-Prinzipien enthalten und umgekehrt. Nur das FAIR-Prinzip *Accessible*, das den Einsatz von standardisierten, frei zugänglichen und universell einsetzbaren Kommunikations-protokollen fordert, findet sich nicht im selben Umfang in den NISO-Prinzipien (vgl. GO FAIR 2023a). Gleichzeitig scheint es in der heutigen Zeit auch eine durch technische Notwendigkeiten geformte Selbstverständlichkeit zu sein, diese Art von Protokollen einzusetzen. Darüber hinaus fehlt das FAIR-Prinzip F3, das fordert, Metadaten sollen die mit Ihnen verknüpften Daten eindeutig und explizit mittels eines Identifikators darstellen (vgl. GO FAIR 2023a), dies ist jedoch in den Empfehlungen für digitale Objekte der NISO (2007:52) enthalten.

Eine wichtige Entwicklung, die sich seit dem Entstehen der NISO-Prinzipien in den FAIR-Prinzipien manifestiert hat, ist die Betonung der Notwendigkeit von Sprachen für die Darstellung von Metadaten und Daten, die eine maschinelle Verarbeitung ermöglichen. Dies wurde in den NISO-Prinzipien zwar bereits erwähnt und in einigen Aspekten vorgeschlagen (vgl. NISO 2007:58–85), jedoch war dies kein so starker Schwerpunkt wie in den FAIR-Prinzipien. Ein wichtiger Unterschied ist auch die Ausrichtung der FAIR-Prinzipien auf Forschungsdaten und deren Metadaten, die impliziert, dass hier Forschende für Forschende Daten und Metadaten schaffen und aufbereiten. Dies findet sich nicht in den NISO-Prinzipien. Zwar können die digitalen Objekte in jeglichen digitalen Sammlungen immer Forschungsobjekte und damit Forschungsdaten werden. Jedoch wird das nicht automatisch vorausgesetzt, wenn diese digitalen Objekte in die digitalen Sammlungen aufgenommen oder dort abgelegt werden. Die NISO-Prinzipien für gute Metadaten und die FAIR-Prinzipien stellen damit zwar verwandte und ähnliche, aber nicht identische Ansprüche an die digitalen Objekte und ihre Metadaten. Die NISO-Prinzipien sind

eine Möglichkeit, Metadaten zu analysieren, die nicht ausschließlich für Objekte erstellt wurden, die Forschungszwecken dienen, sondern auch administrative Aufgaben erfüllen. Sie sollen daher die Basis für die Analyse dieser Arbeit sein und, wo es notwendig und sinnvoll ist, durch Aspekte der FAIR-Prinzipien und weitere Dimensionen ergänzt werden.

Nichtsdestotrotz ist es wichtig festzuhalten, dass eben weil die FAIR-Prinzipien in den NISO-Prinzipien im Wesentlichen enthalten sind, sie auch auf unterschiedliche Arten von digitalen Objekten angewandt werden können, also nicht nur für digitale Objekte aus und für Forschung, Lehre und Studium. Die FAIR-Prinzipien könnten damit, wie die NISO-Prinzipien auch, für unterschiedliche Arten von Gedächtnisinstitutionen, etwa für Museen und Archive relevant sein. Eine genauere Analyse dieser Möglichkeit müsste aber noch durchgeführt werden.

Dieses Kapitel betrachtete das „Framework of Guidance for Building Good Digital Collections“ der *National Information Standards Organization*, seine sechs Metadaten-Prinzipien und die Sammlungen, auf die sie angewandt werden können. Dann wurden die FAIR-Prinzipien vorgestellt und mit den NISO-Prinzipien für gute Metadaten in Beziehung gesetzt. Es wurde festgestellt, dass die FAIR-Prinzipien in Bezug auf die Güte von Metadaten im Wesentlichen bereits in den NISO-Prinzipien für gute Metadaten enthalten sind. Jedoch bieten die NISO-Prinzipien auch die Möglichkeit, sie auf administrative digitale Objekte anzuwenden. Daher sollen sie als Grundlage zur Kategorienbildung für die Analyse der im Rahmen dieser Arbeit durchgeführten Expert*innen-Interviews herangezogen werden und wo nötig ergänzt werden, etwa durch Elemente der FAIR-Prinzipien (s. Anhang 4). Es mag sich die Frage stellen, weshalb die sich ähnelnden und zum Teil ergänzenden NISO- und FAIR-Prinzipien dazu dienen sollen, zu untersuchen, was zu Mappings führen kann und welche Vorteile sie bieten. Die Antwort ist eine Hypothese: Es könnte sein, dass genau diese Prinzipien in einem wellenartigen Prozess Mappings notwendig machen, um sich selbst gerecht zu werden. Dafür spricht auch, dass die wichtigsten Aspekte sich sowohl in den NISO-Prinzipien und den weitaus später entstandenen FAIR-Prinzipien finden. Aus der Ähnlichkeit der beiden Ansätze könnte auch die Notwendigkeit abgeleitet werden, zu untersuchen, ob die Anwendung der FAIR-Prinzipien

in unterschiedlichen Arten von Gedächtnisinstitutionen sinnvoll wäre, die digitale Objekte speichern, welche über Objekte für Forschung, Lehre und Studium hinausgehen.

4. Die Methoden: das Expert*innen-Interview und die qualitative Inhaltsanalyse

Dieses Kapitel legt die Methoden dar, die zur Erhebung und Analyse in dieser Arbeit verwendet werden: zunächst die Methode des Expert*innen-Interviews, die als Erhebungsinstrument für diese Masterarbeit gewählt wurde. Dabei wird das Ziel dieser Methode dargelegt, das im Falle dieser Arbeit in der Informationsgewinnung liegt. Daraufhin werden die einzelnen Schritte vorgestellt, die zur Erstellung eines Interviewleitfadens für das Expert*innen-Interview notwendig sind und in dieser Arbeit angewandt werden. Dazu gehören die Sammlung von Forschungsfragen, die Gruppierung der Forschungsfragen, die Methodenspezifizierung, der Entwurf von Leitfadenfragen, die Differenzierung von Fragetypen für den Leitfaden sowie der Pretest.

Danach wird die Auswertungsmethode der qualitativen Inhaltsanalyse und ihr Ansatz der deduktiven Kategorienanwendung präsentiert. Auch hier werden die einzelnen Schritte betrachtet und in einem Ablaufmodell festgehalten und für diese Arbeit umgesetzt. Dazu gehören die Festlegung des Materials (mit Vorstellung der befragten Expert*innen), die Analyse der Entstehungssituation, die formalen Charakteristika des Materials, Richtung der Analyse und Ablaufmodell der Analyse, theoriegeleitete Festlegung der Kategorien für die Strukturierung und deduktive Kategorienanwendung, Kodierung eines ersten Textteils und Überarbeitung der Kategorien und des Kodierleitfadens, der endgültige Materialdurchgang und die Zuordnung der Kategorien zu Textpassagen sowie die Auswertung.

4.1. Die Erhebungsmethode: das Expert*innen-Interview

Das Expert*innen-Interview ist eine weit verbreitete Methode in den sozialwissenschaftlichen Disziplinen und wird in den unterschiedlichsten Fächern eingesetzt. Diese reichen „von der Pädagogik und Psychologie bis hin zu Politikwissenschaft und Soziologie“ (Bogner u. a. 2014:1). Für Bogner, Littig und Menz (2014) gibt es zwei wichtige Arten

des Expert*innen-Interviews, auf die sie sich für einen pragmatischen Zugang zur Forschungspraxis konzentrieren. Da es Schnittmengen und Parallelen gibt (vgl. Bogner u. a. 2014:2), sollen beide erwähnt werden. Für die erste Variante ist nicht die Erhebung von Fakten, sondern vielmehr […] die Rekonstruktion subjektiver Deutungen und Interpretationen“ wichtig (Bogner u. a. 2014:2). Dieser Ansatz wird auch als „theoriegenerierend“ (Bogner u. a. 2014:2) bezeichnet. Die zweite Variante des Expert*innen-Interviews dient der Informationsgewinnung und wird für diese Arbeit relevant sein (vgl. Bogner, Littig, und Menz 2014: 2, sie verweisen auf Gläser und Laudel 2004). Auch wenn es kritische Stimmen im Hinblick dazu gibt, Expert*innen-Interviews für Informationsgewinnung anzuwenden – Bogner et al. (2014: 3) verweisen auf Lamnek (2005:333) – , spricht einiges auch für

> „das Experteninterview als pragmatische Methode, die man vor allem aus forschungsökonomischen Gründen wählt. Die (plausible) Annahme lautet: Sofern die angefragten Experten ‚Kristallisationspunkte‘ relevanten Insiderwissens sind, ersparen sich die Forscherinnen mit diesen Interviews lange Wege. (Bogner u. a. 2014:2 f.)

Sich lange Wege zu ersparen, muss nicht der einzige Grund sein, der zum Einsatz von Expert*innen-Interviews in der Forschung führt. Vielmehr können Expert*innen auch die oben erwähnten Kristallisationspunkte von Wissen innerhalb einzelner Institutionen sein. Sie können aufgrund der Aktualität des Forschungsinteresses und ausschließlich interner Dokumentation zu bestimmten Fragestellungen verlässlichen Zugang zu Wissen, Prozessen und Entscheidungen bieten. Aus genau diesen Gründen soll das Expert*innen-Interview zur Informationsgewinnung für diese Arbeit eingesetzt werden.

4.1.1. Erstellung des Interviewleitfadens für das Expert*innen-Interview

Für die Erstellung eines Interviewleitfadens für Expert*innen-Interviews gibt es keine Schablone. Abhängig von den Forschungsthemen und dem jeweiligen Forschungsdesign, unterscheidet sich die Vorgehensweise immer wieder (vgl. Bogner u. a. 2014). Bogner, Littig und Menz nennen daher nur einige Schritte, die zu beachten sind und hier für die Struktur dieser Arbeit in der Reihenfolge angepasst wurden. Zunächst wird die Arbeit an den Forschungsfragen betrachtet: mit a) der Sammlung und Systematisierung der For-

schungsfragen und b) der Gruppierung der Forschungsfragen. Daran schließt c) die Methodenspezifizierung an. Daraufhin folgt die Arbeit an den Leitfadenfragen mit d) dem Entwurf der Leitfadenfragen, e) der Differenzierung von Fragetypen für den Leitfaden und daran schließt f) der Pretest an, um die eigentlichen Interviews vorzubereiten. Diese unterschiedlichen Schritte sollen nun beschrieben werden.

a) Die Sammlung und Systematisierung der Forschungsfragen

In einem ersten Schritt geht es darum, alle vorhandenen Forschungsfragen zusammenzutragen und etwa mittels einer Mind-Map zu strukturieren und sie gleichzeitig strukturell und inhaltlich zu formen und zu reduzieren, um am Ende möglichst klare Forschungsfragen zu haben, die durch die Expert*innen-Interviews beantwortet werden sollen. Ausgangspunkt dafür können die Forschungsfragen aus dem Exposé oder einer wissenschaftlichen Abschlussarbeit sein (vgl. Bogner u. a. 2014:33). Für die vorliegende Arbeit werden die Forschungsfragen aus dem Exposé herangezogen und als Basis für die Weiterentwicklung und Strukturierung verwendet. Diese lauten:

- Was kann Mappings zwischen unterschiedlichen Metadatenformaten in digitalen Repositorien notwendig machen?
- Welche Vorteile können sich durch Mappings in andere Metadatenformate ergeben?

b) Gruppierung

Wenn die Forschungsfragen gesammelt, strukturiert, untersucht und ausgewählt wurden, werden sie größeren thematischen Blöcken zugeordnet, die Oberbegriffe mit jeweils untergeordneten Themen und Unterfragen (vgl. Bogner u. a. 2014:33).

c) Methodenspezifizierung

Die Methodenspezifizierung betrachtet, ob es innerhalb der Sammlung von Forschungsfragen Fragen gibt, die besser durch eine andere Methode als das Expert*innen-Interview

beantwortet werden könnten. Außerdem wird hier ausgewählt, welche spezifische Form des Expert*innen-Interviews verwendet werden soll (vgl. Bogner u. a. 2014:33).

Da die Forschungsfragen dieser Arbeit durch das Expert*innen-Interview beantwortet werden können, wird hier ein zweiter Aspekt relevant, denn an diesem Punkt „wird […] entschieden, für welche Formen des Experteninterviews wir uns entscheiden. In welchem Verhältnis […] ‚Information' (Fakten, Prozesse) und ‚Theoriegenerierung' [stehen]" (Bogner u. a. 2014:33). Um diese Frage klären zu können, ist es notwendig, die Arten von Wissen zu betrachten, die durch das Expert*innen-Interview erhoben werden können: Abhängig von der Art des Wissens, auf das zugegriffen werden soll, unterscheidet sich die Interviewform, die für ein Expert*innen-Interview angewandt wird. Es gibt drei unterschiedliche Arten von Wissen im Bereich der Expert*innen-Interviews: das technische Wissen, das Prozesswissen und das Deutungswissen[8] (vgl. Bogner u. a. 2014:17). Wichtig für diese Arbeit sind die ersten beiden Arten des Wissens: das technische Wissen und das Prozesswissen. Daher sollen diese beiden Arten des Wissens nun vorgestellt werden.

> „Unter ***technischem Wissen*** verstehen wir Daten, Fakten, ‚sachdienliche Informationen', Tatsachen. Die Befragten verfügen über privilegierte Zugänge zu bestimmten Informationen, über die wir als Forscher nicht verfügen. Um an diese Informationen zu gelangen, führt kein Weg daran vorbei, ihn zu fragen." (Bogner u. a. 2014:17f. H.i.O.)

Es ist deshalb wesentlich Expert*innen auszuwählen, die genau über diese privilegierten Zugänge verfügen und diese Informationen mittels eines Interviews zu erschließen. Wichtig zu bemerken ist, dass diese Art von Wissen „vom Träger des Wissens abstrahiert werden [kann], es ist gewissermaßen personenunabhängig. Technisches Wissen ist Wissen über Dinge oder Zusammenhänge, von dem wir annehmen, dass es objektiv ist."(Bogner u. a. 2014:17). Gleichzeitig ist auch diese Perspektive Resultat eines konstruktivistischen Moments, in dem die Expert*innen als „Überbringer von Informationen [konzeptualisiert werden]" (Bogner u. a. 2014:17) und deshalb müssen auch die so gewonnenen Informationen sorgfältig analysiert werden. Die Zuverlässigkeit und Richtigkeit der durch

[8] Das Deutungswissen bezieht sich stärker auf subjektive Aspekte der Wahrnehmung: „die subjektiven Relevanzen, Sichtweisen, Interpretationen, Deutungen, Sinnentwürfe Erklärungsmuster" (Bogner u. a. 2014) und ist daher für diese Arbeit nicht relevant.

die Expert*innen-Interviews gewonnenen Informationen kann selbst bei sorgfältigster und gewissenhaftester Vorbereitung durch beide Seiten, sowohl die Expert*innen als auch die Forschenden, nicht garantiert werden. Denn „die Befragte könnte sich irren, nicht gut informiert sein, sich nicht auskennen usw. – sie ist eine potenzielle Fehlerquelle." (Bogner u. a. 2014:17)

Für diese Arbeit ist Methode des Expert*innen-Interviews dennoch eine gute und gangbare Methode, da es zum Prozess des Metadaten-Mappings des von der Universitätsbibliothek der Universität Wien in PHAIDRA verwendeten Formats UWMETADATA in ein weiteres Format (außer Fiala & Huggle 2019) keine Publikationen gibt. Daher sind die Expert*innen-Interviews eine wichtige, wenn nicht sogar die einzige Möglichkeit, an das relevante Wissen zu kommen. Denn:

> „Experteninterviews, die auf technisches Wissen zielen, [können] eine sinnvolle Abkürzungsstrategie sein: Die Informationen sind anderweitig schwer zu besorgen, oder ihre Beschaffung dauert zu lange, oder die Informationen sind in der gewünschten Form gar nicht zugänglich.

Die zweite Art des Wissens, die für diese Arbeit relevant ist, ist das Prozesswissen der zu befragenden Expert*innen:

> „es umfasst Einsicht in Handlungsabläufe, Interaktionen, organisationale Konstellationen, Ereignisse usw., in die die Befragten involviert sind oder waren. Die Befragten haben aufgrund von Erfahrungen, aufgrund einer persönlichen Nähe zu Ereignissen ein bestimmtes Wissen." (Bogner u. a. 2014:17)

Die für diese Arbeit ausgewählten Expert*innen sollten genau diese Erfahrungen haben, da sie an der Universitätsbibliothek der Universität Wien, also der Institution, die diese Analyse betrifft, arbeiten und den gesamten Prozess des Mappings aus nächster Nähe beobachtet und auch daran mitgewirkt habe. Darüber hinaus sind sie auch in zukünftige Entscheidungen in diesem Bereich eingebunden. Da diese Prozesse bisher nicht an die (bibliothekarische) Öffentlichkeit kommuniziert wurden und die Kommunikation vor allem intern passierte, sind die gewählten Expert*innen eine wichtige Quelle für dieses Wissen. Betrachtet man technisches Wissen und Projektwissen, handelt es sich um Wissen, das relativ leicht zugänglich ist, denn es ist für die

> „Befragungspersonen selbst jederzeit reflexiv verfügbar ist. Wir können also dieses Wissen mehr oder weniger direkt abfragen, und wir benötigen keine sonderlich ausgefeilten hermeneutischen

> Techniken, etwa, weil wir verdecktes Wissen ans Tageslicht befördern wollen.“ (Bogner u. a. 2014:24)

Um an dieses Wissen zu kommen, das als technisches Wissen und Prozesswissen vorhanden ist, empfiehlt sich das systematisierende Expert*innen-Interview (vgl. Bogner u. a. 2014:24). Ziel dieser Interviewform liegt in einer „möglichst weitgehenden und umfassenden Erhebung des Sachwissens der Experten bezüglich des Forschungsthemas. Das Interview dient der systematischen Informationsgewinnung“ (Bogner u. a. 2014:24).
Für das Interview ist ein ausdifferenzierter Leitfaden zu erarbeiten, um strukturiert alle Fragen zu stellen und so alles Wissen zu erheben, das für die Forschenden notwendig ist (vgl. Bogner u. a. 2014:24). Um das zu erreichen, ist zunächst der nachfolgende Aspekt im Forschungsdesign notwendig.

d) Entwurf von Leitfadenfragen

Der nächste wichtige Schritt im Forschungsdesign für Bogner, Littig und Menz ist folgender: „Es geht jetzt […] um die Formulierung von Interviewfragen, also solchen Fragen, die dann auch im eigentlichen Erhebungsgespräch tatsächlich gestellt werden könnten.“ (Bogner u. a. 2014:33f.). Dabei soll vor allem für die Expert*innen im Gespräch Raum geschaffen werden, damit sie die Möglichkeit und den Wunsch haben, ihr Wissen zu kommunizieren und davon zu berichten. Sie sollen sich dabei auf diejenigen Themen konzentrieren, die für die jeweilige Forschungsarbeit relevant sind. Gleichzeitig ist es möglich, dass sich die Fragestellung der Arbeit durch die Interaktion mit den Expert*innen weiterentwickeln kann (vgl. Bogner u. a. 2014:32).

Das Interview regt also

> „Bewertungen, Schilderungen, Erzählungen zu bestimmten Themen an – wir stellen den Befragten nicht einfach unsere Forschungsfrage […]. Das „Scharnier“ zwischen Forschungsfragen und Leitfaden- bzw. Interviewfragen ist der thematische Fokus, nicht die formulierte Frage. Die Logik der Forschungsfragen unterscheidet sich grundsätzlich von der Logik der Interviewfragen: Erstere sind formuliert in Hinblick auf theoretische Annahmen, letztere sind formuliert in Hinblick auf den Wissens- und Erfahrungshorizont der Befragten.“ (Bogner u. a. 2014:33)

Für die Interviews müssen Fragen gesammelt und entwickelt werden, um daraus einen Interviewleitfaden zu erstellen. Expert*innen-Interviews sind teilstrukturierte Interviews.

Die Leitfäden dienen daher dazu, einerseits für den Forschenden das zu untersuchende Fachgebiet zu gliedern, gleichzeitig unterstützen sie das Expert*innen-Interview (vgl. Bogner u. a. 2014:28).

e) Differenzierung von Fragetypen

Nachdem die Leitfadenfragen entworfen wurden, werden sie wieder geordnet und in Hauptfragen und Unterfragen sowie Nachfragen unterteilt. Daraufhin werden sie strukturiert und in eine Reihenfolge gebracht. Diese kann sich während des Interviews auch verändern (vgl. Bogner u. a. 2014:34).

Ein fertiger Interviewleitfaden für ein Interview von ein bis zwei Stunden besteht normalerweise

> „aus verschiedenen Themenblöcken […] etwa drei bis maximal acht. Zu jedem Themenblock sind jeweils ein bis drei Hauptfragen notiert, die als zentrale Gesprächsanreize zum jeweiligen Themenabschnitt dienen. Ergänzend kommen verschiedene abhängige Fragen hinzu, die der weiteren Detaillierung dienen […]." (Bogner u. a. 2014:28f.)

Alle Hauptfragen müssen während der Expert*innen-Interviews gestellt werden, die thematisch und im Hinblick auf ihre Relevanz vergleichbar sind. Die komplementären Unterfragen, die mehr ins Detail gehen, können im Interview herangezogen werden, wenn während der Beantwortung der Hauptfragen noch Aspekte offengeblieben sind (Bogner u. a. 2014:29)

Generell gilt, dass auch die Interview-Leitfäden gut strukturiert und graphisch aufbereitet sein sollten, um während des Interviews möglichst wenig Kapazitäten für die Orientierung darin verwenden zu müssen. Was die Länge betrifft, sollten Leitfäden nicht zu lang sein, eine Länge von 1-6 Seiten wird im Allgemeinen als ausreichend erachtet, wobei Fragen im Verlauf des Gesprächs auch anders formuliert werden können, als geplant, oder auch ausgelassen werden können, da sie bereits erschöpfend besprochen wurden. Auch der Gesprächsfluss kann Auswirkungen auf die Reihenfolge der Fragen haben. Manchmal wird so früh ein Thema besprochen, das im Interviewleitfaden erst später vorhergesehen war. Daher sollten die einzelnen Themenblöcke in sich geschlossen sein, sodass sie an eine andere Stelle im Interview verschoben werden können, wenn dies für die

Erhaltung des Gesprächsflusses notwendig sein sollte. Der Gesprächsfluss kann es teilweise auch notwendig machen, spontan Nachfragen zu stellen, die nicht im Interviewleitfaden vorhergesehen waren (vgl. Bogner u. a. 2014:29f.).

Sollten sich die Expert*innen in dem für die Forschungsarbeit relevanten Wissen extrem unterscheiden, kann es auch notwendig sein, die Interview-Leitfäden an die jeweilige*n Interview-Partner*in anzupassen und eine abgeänderte Fassung zu erstellen, um diesen Unterschieden gerecht zu werden. Auch wenn der Kenntnisstand der Forschenden sich nach einem Interview verändert hat, kann der Leitfaden angepasst werden (vgl. Bogner u. a. 2014:30).

Wenn es um die Übermittlung von Interview-Leitfäden vor dem jeweiligen Gespräch geht, kann es für die für diese Arbeit relevanten systematisierenden Expert*innen-Interviews von Nutzen sein, denn

> „wenn wir auf die Erhebung von Detailinformationen abzielen, die die Befragten im Gespräch spontan vermutlich nicht präsent haben, kann das Zurechtlegen und Aufbereiten von Informationen vor dem eigentlichen Interview von Seiten der Befragten sinnvoll sein. Häufig ist aber davon auszugehen, dass die Expertinnen sich auf das Gespräch nicht vorbereiten müssen.“ (Bogner u. a. 2014:30).

Da sich die Fragestellung dieser Arbeit auf ein stark technisches Thema, das über Jahre zurückreicht, konzentriert, wird der Interviewleitfaden den Expert*innen vor dem Interview zugesandt.

f) Pretest

Falls es möglich ist, eine*n weitere*n Expert*in zu gewinnen, sollte auch ein Pretest für den Leitfaden durchgeführt werden, um herauszufinden, ob er praktikabel ist und sich zeitlich umsetzen lässt. Auch sollte das Feedback diese*r Expert*in eingeholt werden, um festzustellen, ob eventuell Fragen hinzugefügt werden können, fehlen oder unverständlich sind (vgl. Bogner u. a. 2014:34).

Für diese Arbeit wird kein Pretest durchgeführt, da es nur eine sehr begrenzte Anzahl von Expert*innen an Bibliothek der Universität Wien zum Thema dieser Arbeit gibt und diese alle auch befragt werden sollen.

g) Dokumentation

Ein weiterer wichtiger Aspekt, der berücksichtig und geplant werden muss, ist die Dokumentation der Interviews. Sie findet, wie von Bogner, Littig und Menz (vgl. 2014:40) empfohlen, in Form einer Tonaufnahme statt (sofern die Expert*innen dies nicht ablehnen sollten. Andernfalls wird es so gut wie möglich handschriftlich notiert).

Da es sich bei den Expert*innen-Interviews dieser Arbeit um systematisierende Interviews handelt, bei denen es weniger um die Erfassung von Deutungswissen, sondern vielmehr um das Erfassen von spezifischen Informationen geht, können im Notfall Interviewprotokolle für die Dokumentation ausreichen. Denn im Zentrum steht hier nicht der detaillierte Wortlaut, der von den Expert*innen verwendet wurde und noch weiterer Analyse und Interpretation bedarf, sondern vor allem der Informationsgehalt (vgl. Bogner u. a. 2014:41). Dennoch soll versucht werden, die Zustimmung der Expert*innen zu einer Tonaufnahme der Interviews zu gewinnen.

Wenn der Leitfaden ausgearbeitet ist und alle Vorbereitungen abgeschlossen, folgt die Auswahl der Expert*innen für das Interview und dieser Prozess soll nun beschrieben werden.

4.1.2. Auswahl der Expert*innen/ Sampling

Die Auswahl der Expert*innen wird einerseits durch das jeweilige Forschungsinteresse bestimmt, also durch den fachlichen Kontext und andererseits durch die Rolle und Funktion, die sie innerhalb der Gesellschaft haben. Beide Aspekte gemeinsam konstituieren Expert*innen (vgl. Bogner u. a. 2014:11): Forschende kontaktieren Menschen vor dem Hintergrund eines bestimmten Forschungsinteresse, um Interviews mit ihnen durchzuführen. Durch die Auswahl und die Aufgabe, die sie im Hinblick auf das Forschungsinteresse haben, werden diese Menschen zu Expert*innen (vgl. Bogner u. a. 2014). Gleichzeitig ist dieser Expert*innen-Status in der gesellschaftlichen Position der zu Interviewenden begründet:

> „Schließlich greifen wir bei der Expertenauswahl in der Regel auf Personen zurück, die gemeinhin als Experten gelten; wir rekurrieren also immer auch auf gesellschaftliche Konventionen. Im Regelfall gelten Experten als Personen, die in herausgehobenen sozialen Positionen und in solchen Kontexten handeln, die sie als Experten kenntlich machen; sie sitzen beispielsweise in einer Expertenkommission, einem Beratungsgremium, haben einen Professorentitel und Ähnliches mehr. Kurz gesagt: Experten sind Angehörige der „Funktionselite" (Meuser und Nagel 1994, S. 181)." (Bogner u. a. 2014:11)

Aus dieser Aussage leiten Bogner, Littig und Menz eine Definition von Expert*innen her:

> „Experten lassen sich als Personen verstehen, die sich – ausgehend von einem spezifischen Praxis- oder Erfahrungswissen, das sich auf einen klar begrenzbaren Problemkreis bezieht – die Möglichkeit geschaffen haben, mit ihren Deutungen das konkrete Handlungsfeld sinnhaft und handlungsleitend für Andere zu strukturieren." (Bogner u. a. 2014:13)

Expert*innen sind daher Menschen, die in einem klar definierten Bereich durch Praxis oder Erfahrung Kenntnisse und Wissen erlangt haben, die es ihnen ermöglichen, den jeweiligen Bereich und die darin auftretenden Probleme für andere Akteure so zu gestalten, sodass diese sinnvolle Handlungen setzen können (vgl. Bogner u. a. 2014:13). Expert*innen haben damit Gestaltungsmacht, die über die eigene Institution hinausreichen kann (vgl. Bogner u. a. 2014:13). Darüber hinaus beschränken sie sich nicht auf ihren eigenen Bereich, sondern besitzen die „Fähigkeit, Verbindungen zu anderen Wissensbeständen und Wissensformen herzustellen und die Relevanz des eigenen Wissens zu reflektieren" (Hitzler 1994:21 f. zit. in Bogner u. a. 2014:15). Dieser Aspekt ist besonders wichtig im Hinblick auf die Zusammenarbeit und Wirkung jenseits des eigenen Fachbereichs. Gleichzeitig ermöglicht die Fähigkeit Verbindungen zu anderen Wissensbeständen herzustellen, an der Schnittstelle unterschiedlicher Fachbereiche zu arbeiten, wie es in der Gegenwart immer notwendiger wird und auch die Fragestellungen, die dieser Arbeit zugrunde liegen, betrifft.

> „Schütze (1992) paraphrasierend kann man abschließend festhalten: Experten besitzen die Möglichkeit zur (zumindest partiellen) Durchsetzung ihrer Orientierungen. Experten zeichnen sich dadurch aus, dass sie maßgeblich bestimmen, aus welcher Perspektive und mithilfe welcher Begrifflichkeiten in der Gesellschaft über bestimmte Probleme nachgedacht wird. (Bogner u. a. 2014:15)"

Experten können daher Entscheidungsmöglichkeiten formen, die zukünftige Schritte und Prozesse in der Gesellschaft, aber auch innerhalb einer Institution beeinflussen können, wie es relevant für diese Arbeit ist. Dieses in den Expert*innen liegende Potential ist deshalb Ausgangspunkt dieser Arbeit.

4.2. Die Auswertungsmethode: die qualitative Inhaltsanalyse

Die qualitative Inhaltsanalyse nach Mayring (1983/2022) ist eine Methode, die es ermöglicht, komplexes Material, das in sprachlicher Form vorliegt, regelgeleitet auszuwerten und den Prozess der Analyse zu strukturieren, zu systematisieren und nachvollziehbar zu machen (vgl. Mayring 2022:49). Sie soll als Auswertungsmethode für diese Arbeit nun in ihren einzelnen Schritten vorgestellt werden.

Ausgangspunkt für die qualitative Inhaltsanalyse als Auswertungsmethode ist sprachliches Material. Dieses Material muss genau beschrieben werden, bevor es untersucht wird. Zumindest drei Schritte sind dafür notwendig, die nun unter a), b) und c) dargestellt werden sollen (vgl. Mayring 2022:53), bevor dann die restlichen Analyseschritte erarbeitet werden.

a) Festlegung des Materials

In jeder inhaltsanalytischen Arbeit muss zunächst definiert werden, welche Texte Gegenstand der Analyse sein sollen. Dieser Textkorpus sollte möglichst keinen Änderungen während der Auswertung unterworfen werden. Dabei sind neben der Definition des Corpus auch Stichprobenumfang einschließlich repräsentativer und ökonomischer Überlegungen, sowie Stichprobenziehung nach einem festgelegten Modell festzulegen (vgl. Mayring 2022:53f.).

Bei den für diese Arbeit vorliegenden Interviews handelt es sich um drei Expert*innen-Interviews, die anhand eines teilstrukturierten Interviewleitfadens durchgeführt wurden. Inhaltlich beschäftigen sich die Interviews mit Metadaten und Metadaten-Strukturen, die für das digitale Repositorium der Universität Wien geschaffen und verwendet wurden und in Zukunft verwendet werden. Bei der Auswahl der Expert*innen zum Zweck der Informationsgewinnung standen Rolle und Kenntnisse der Expert*innen im Mittelpunkt, die einen ökonomischen Zugang zu den relevanten Informationen rund um das Themengebiet ermöglichten. Daher kann die Auswahl nicht als repräsentativ gewertet werden.

Die Expert*innen sind:

Expertin 1: Susanne Blumesberger

Susanne Blumesberger ist Leiterin der Abteilung Repositorienmanagement PHAIDRA – Services an der Universitätsbibliothek der Universität Wien und wirkte an der Entwicklung und dem Aufbau der Repositorium PHAIDRA an der Universität mit. Blumesberger verfasste zahlreiche Publikationen zur Entwicklung des Langzeitarchivierungs-Repositorium PHAIDRA an der Universität Wien, etwa Blumesberger (2009; 2010a; 2010b). Außerdem leitete Blumesberger im Projekt e-Infrastructures Austria, das in Kapitel 1 vorgestellt wurde, den Cluster I: „Metadatenkomplex aus nicht-technischer und technischer Sicht". Blumesberger leitet darüber hinaus im Nachfolgeprojekt e-Infrastructures Austria Plus[9], das Arbeitspaket 4.5. Metadaten. Sie publiziert u.a. zum Thema Metadaten und digitale Repositorien für die Langzeitarchivierung, etwa in Blumesberger (2015; 2017; 2019). https://orcid.org/0000-0001-9018-623X

Experte 2: Raman Ganguly

Raman Ganguly ist Leiter der Abteilung IT-Support for Research des Zentralen Informatikdiensts der Universität Wien und wirkte an der technischen Entwicklung von PHAIDRA einschließlich der Arbeit mit Metadaten und kontrollierten Vokabularen in PHAIDRA mit, wie etwa in Subitras et al. (2017) zeigt. Außerdem wirkte er generell im Bereich der Metadaten-Organisation in digitalen Repositorien, die für die Langzeitarchivierung eingesetzt werden, etwa Kopácsi, Hudak und Ganguly (2017). Außerdem leitete Ganguly im vom österreichischen Bundesministerium für Wissenschaft, Forschung und

[9] Das Projekt e-Infrastructures Austria Plus wurde wie das Projekt e-Infrastructures Austria auch vom österreichischen Bundesministerium für Wissenschaft, Forschung und Wirtschaft gefördert. Es setzte sich zwischen 2017 und 2019 vertieft mit den Veränderungen im Bereich Infrastruktur auseinander, die die Entwicklung von Wissenschaft und Forschung hin zu Open Science notwendig machen. Diese Entwicklung sollte durch den Einsatz von digitaler Infrastruktur vorangetrieben werden. Die (aktuelle) Form wissenschaftlich zu arbeiten, wird auch als e-Science bezeichnet (vgl. Haselwanter 2019:4).

Wirtschaft (BMWFW) geförderten Projekt e-Infrastructures Austria den Cluster D: Aufbau Infrastruktur. Ganguly wirkte auch am Folgeprojekt e-Infrastructures Austria Plus mit, hier im Arbeitspaket 4.7. GO FAIR. https://orcid.org/0000-0002-9837-0047

Experte 3: Rastislav Hudak

Rastislav Hudak ist Mitarbeiter der Abteilung IT-Support for Research des Zentralen Informatikdienst der Universität Wien und hat sich aus technischer Perspektive mit den Mappings des Formats UWMETADATA beschäftigt. Außerdem befasst Hudak sich, wie Ganguly auch, mit der Verknüpfung von Metadaten und kontrollierten Vokabularen in PHAIDRA in Kopácsi, Hudak und Ganguly (2017) und ebenso mit der Organisation von Metadaten in digitalen Repositorien für Langzeitarchivierung, so in Subitras et al. (2017). Außerdem wirkte er ebenfalls am Projekt e-Infrastructures Austria mit, etwa im Cluster I: Metadaten Projekt e-Infrastructures Austria bei der Erstellung eines Jobprofils für Metadatenmanager*innen (Blumesberger, Gründhammer, u. a. 2016) oder bei der Wahl und Implementierung eines Klassifikationsservers für ein digitales Repositorium für die Langzeitarchivierung (Kopácsi u. a. 2017).

Die drei Expert*innen arbeiten alle an der Universität Wien aus ihren spezifischen Positionen heraus an der Entwicklung von PHAIDRA und sind– in unterschiedlichem Ausmaß - an der Arbeit mit den unterschiedlichen Metadatenformaten beteiligt gewesen. Eine Reihe von weiteren möglichen Interview-Teilnehmer*innen, die an der Universitätsbibliothek der Universität Wien rund um das Thema Metadaten in PHAIDRA arbeiteten, wurden nach einer Empfehlung von Sonja Fiala per Mail kontaktiert. Die meisten kontaktierten Personen konnten oder wollten nicht an den Interviews teilnehmen, schlugen meist aber weitere mögliche Expert*innen vor. Die hier Versammelten kristallisierten sich daher aus den Vorschlägen von Sonja Fiala und denen der Kontaktpersonen als Expert*innen für das Themengebiet heraus und stimmten schließlich den Interviews zu.

b) Analyse der Entstehungssituation

In einem zweiten Schritt der qualitativen Inhaltsanalyse muss genau festgehalten werden, wie das sprachliche Material entstanden ist und unter welchen Umständen. Dazu gehören Informationen zu den an der Texterstellung Beteiligten, ihrem Handlungshintergrund, in manchen Fällen Überlegungen dazu, für welches Zielpublikum die Texte entstanden sind, außerdem die jeweilige Entstehungssituation und, sofern relevant, der soziokulturelle Hintergrund (vgl. Mayring 2022:54).

Im Falle dieser Arbeit basierte die Teilnahme der Expert*innen an den Interviews auf Freiwilligkeit. Bei den Interviews handelt es sich um teilstrukturierte Interviews, die anhand eines Leitfadens geführt wurden, wobei die Reihenfolge und Formulierung der Fragen dem Gesprächsverlauf angepasst werden kann. Die Fragen waren teilweise als offene Fragen und teilweise als geschlossenen Fragen gestaltet. Wobei letztere immer als äö auf ein Thema und Gesprächsanreiz dienten und durch offene Fragen ergänzt wurden. Der Interviewleitfaden wurde den Expert*innen vor dem Interview zugesandt. Die Interviews wurden von der Autorin in den Räumen der Wirtschaftsuniversität Wien durchgeführt. Das Interview mit Raman Ganguly fand am 20.12.2023 statt. Das Interview mit Susanne Blumesberger fand am 15.01.2024 und das mit Rastislav Hudak am 30.01.2024 statt. Da einige wenige Fragen nach den Interviews mit Susanne Blumesberger und Rastislav Hudak noch offengeblieben waren, wurden diese noch per Mail am 05.03.2024 an die Expert*innen gesandt und von Susanne Blumesberger am 13.03.2024 und von Rastislav Hudak am 15.03.2024 beantwortet.

c) Formale Charakteristika des Materials

Der dritte Schritt bei der Bestimmung des Ausgangsmaterials besteht darin, die Form des Textmaterials festzuhalten. Auswertungen mittels der qualitativen Inhaltsanalyse geschehen meist auf Basis eines schriftlichen Texts. Da Interviews häufig aufgenommen werden, um dann durch Transkription in einen geschriebenen Text umgewandelt zu werden, müssen auch die Transkriptionsregeln bekanntgegeben werden (vgl. Mayring 2022:54).

Das Material, das dieser Arbeit zugrunde liegt, basiert auf Interviews, die mit einem Mobiltelefon aufgenommen und dann zunächst automatisch mit Word für Microsoft 365 transkribiert und daraufhin mehrfach unter Verwendung der Aufnahmen kontrolliert und (entsprechend der gewählten Transkriptionsregeln) korrigiert wurden. Das Transkript soll vor allem die Analyse inhaltlich-semantischer Aspekte der Interviews möglich machen soll. Daher werden Transkriptionsregeln verwendet, die leichte Lesbarkeit der transkribierten Interviews ermöglichen, sich leicht angeeignet und umgesetzt werden können (vgl. Dresing & Pehl 2018:17). Dresing und Pehl (2018) verweisen auf Kallmeyer und Schütze (1976) und Hoffmann-Riem (1984), die solche Ansätze erarbeitet und im Falle von Kuckartz et al. (2008) ergänzt und für die computergestützte Arbeit weiterentwickelt haben. Dresing und Pehl bauen auf diesen Kenntnissen auf und entwickeln einen inhaltlich-semantischen Transkriptionsansatz, der die bisherigen Ansätze überarbeitet und ergänzt (vgl. Dresing & Pehl 2018:17). Die Transkriptionsregeln lauten:

> „1. Es wird wörtlich transkribiert, also nicht lautsprachlich oder zusammenfassend.
>
> 2. Wortverschleifungen werden an das Schriftdeutsch angenähert. „So‘n Buch“ wird zu „so ein Buch“ und „hamma“ wird zu „haben wir“. Die Satzform wird beibehalten, auch wenn sie syntaktische Fehler beinhaltet, z.B.: „Bin ich nach Kaufhaus gegangen.“
>
> 3. Dialekte werden möglichst wortgenau ins Hochdeutsche übersetzt. Wenn keine eindeutige Übersetzung möglich ist, wird der Dialekt beibehalten, z.B.: „Ich gehe heuer auf das Oktoberfest“.
>
> 4. Umgangssprachliche Partikeln wie „gell, gelle, ne“ werden transkribiert.
> 5. Stottern wird geglättet bzw. ausgelassen, abgebrochene Wörter werden ignoriert. Wortdoppelungen werden nur erfasst, wenn sie als Stilmittel zur Betonung genutzt werden: „Das ist mir sehr, sehr wichtig.“
>
> 6. Halbsätze, denen die Vollendung fehlt, werden mit dem Abbruchzeichen „/“ gekennzeichnet.
>
> 7. Interpunktion wird zugunsten der Lesbarkeit geglättet, das heißt, bei kurzem Senken der Stimme oder nicht eindeutiger Betonung wird eher ein Punkt als ein Komma gesetzt. Sinneinheiten sollten beibehalten werden.
>
> 8. Rezeptionssignale wie „hm, aha, ja, genau“, die den Redefluss der anderen Person nicht unterbrechen, werden nicht transkribiert. Sie werden dann transkribiert, wenn sie als direkte Antwort auf eine Frage genannt werden.
> Pausen ab ca. 3 Sekunden werden durch (…) markiert.
>
> 10. Besonders betonte Wörter oder Äußerungen werden durch VERSALIEN gekennzeichnet.
>
> 11. Jeder Sprecherbeitrag erhält eigene Absätze. Zwischen den Sprechern gibt es eine freie, leere Zeile. Auch kurze Einwürfe werden in einem separaten Absatz transkribiert. Mindestens am Ende eines Absatzes werden Zeitmarken eingefügt.

> 12. Emotionale nonverbale Äußerungen der befragten Person und des Interviewers, welche die Aussage unterstützen oder verdeutlichen (wie lachen oder seufzen), werden beim Einsatz in Klammern notiert.
> 13. Unverständliche Wörter werden mit „(unv.)“ gekennzeichnet. Längere unverständliche Passagen werden möglichst mit der Ursache versehen: „(unv., Mikrofon rauscht)“. Vermutet man einen Wortlaut, wird die Passage mit einem Fragezeichen in Klammern gesetzt, z.B. „(Axt?)“. Unverständliche Stellen werden mit einer Zeitmarke versehen, wenn innerhalb von einer Minute keine weitere Zeitmarke gesetzt ist.
>
> 14. Die interviewende Person wird durch ein „I:“, die befragte Person durch ein „B:“ gekennzeichnet. Bei mehreren Interviewpartnern (z.B. Gruppendiskussion) wird dem Kürzel „B“ eine entsprechende Kennnummer oder ein Name zugeordnet („B1:“, „Peter:“).
>
> 15. Das Transkript wird als Rich Text Format (RTF-Datei) gespeichert. Die Benennung der Datei erfolgt entsprechend dem Mediendateinamen (ohne Endung wav, mp3), beispielsweise: Interview_04022011.rtf oder Interview_schmitt.rtf.“ (Dresing & Pehl 2018:21 H.i.O.)

Diese 15 Vorgaben sollen für die Transkription der Interviews dieser Arbeit ergänzt und leicht modifiziert werden. Da das Interesse dieser Arbeit vor allem auf dem Aspekt der Informationsgewinnung liegt wird Regel 2 leicht verändert und lautet für die Interviews dieser Arbeit:

> 2. Wortverschleifungen werden an das Schriftdeutsch angenähert. „So‘n Buch“ wird zu „so ein Buch“ und „hamma“ wird zu „haben wir“. Die Satzform wird beibehalten, allerdings werden syntaktische und grammatikalische Fehler korrigiert, um die Lesbarkeit zu verbessern.

Regel 5 wird auch erweitert und lautet für diese Arbeit:

> 5. Stottern wird geglättet bzw. ausgelassen, abgebrochene Wörter werden ignoriert. Wortdoppelungen werden nur erfasst, wenn sie als Stilmittel zur Betonung genutzt werden: „Das ist mir sehr, sehr wichtig.“ Zusätzlich werden Füllwörter ausgelassen.

Diese beiden zusätzlichen Regeln sollen eine Konzentration auf die inhaltlich-semantischen Aspekte der Interviews ermöglichen und die Lesbarkeit erleichtern.

d) Richtung der Analyse

Nachdem das Material, das der Analyse zugrunde liegt, festgelegt und beschrieben wurde, wir in einem weiteren Schritt festgehalten, in welche Richtung die Analyse gehen soll. Dafür ist eine konkrete Fragestellung notwendig. (vgl. Mayring 2022:57):

Da sich auf Grundlage eines Textkorpuses zahlreiche unterschiedliche Analyseperspektiven entwickeln können, ist es wichtig im Voraus festzulegen, worauf die Analyse abzielt. Ausgangspunkt für diese Überlegungen kann ein inhaltsanalytisches Kommunikationsmodell sein. Dieses hält die unterschiedlichen für die Kommunikation relevanten Aspekte fest. Aus der Vielzahl der Aspekte können die für die jeweilige Analyse relevanten ausgewählt werden (vgl. Mayring 2022:58).

Ziel der Analyse dieser Arbeit nach dem inhaltsanalytischen Kommunikationsmodell (vgl. Mayring 2022:58 Abb. 7) ist es vor allem, den kognitiven Hintergrund in Form des Wissenshintergrunds der befragten Expert*innen herauszuarbeiten. Außerdem sollte auch der Handlungshintergrund, in Form von Intentionen und Plänen, sowie bisherigen, auf den Forschungsgegenstand bezogenen Handlungen beleuchtet werden.

e) Theoriegeleitete Differenzierung der Fragestellung

Die qualitative Inhaltsanalyse besteht auf eine „Theoriegeleitetheit der Interpretation" (Mayring 2022:58). Dafür notwendig ist „die Analyse einer präzisen theoretisch begründeten inhaltlichen Fragestellung" (Mayring 2022:58), wie sie für diese Arbeit mittels des Forschungsstands entwickelt wurde. Sie lauten:

- Was kann Mappings zwischen unterschiedlichen Metadatenformaten in digitalen Repositorien notwendig machen?
- Welche Vorteile können sich durch Mappings in andere Metadatenformate ergeben?

f) Ablaufmodell der Analyse

Um Intersubjektivität und Nachvollziehbarkeit der qualitativen Inhaltsanalyse sicherzustellen, muss ein Ablaufmodell der Analyse entwickelt und die eingesetzten Analysetechniken festgehalten werden (vgl. Mayring 2022:60). Beim Ablaufmodell (Abb.1) für diese Arbeit handelt es sich um eine Kombination des allgemeinen Ablaufmodells (vgl. Mayring 2022:61) mit dem Ablaufmodell der deduktiven Kategorienanwendung (vgl. May-

ring 2022:97). Die ersten fünf Schritte dieses Modells bestehen aus den bisher erarbeiteten Analyseschritten. Diesen dienten dazu, das Material, das analysiert werden soll, und die Aspekte, die daraus mittels einer konkreten Fragestellung erarbeitet werden sollen, festzulegen (vgl. Mayring 2022:70). Die weiteren Schritte des Ablaufmodells werden nun schrittweise vorgestellt.

g) Bestimmung der Analyseeinheiten: Kodier- ,Kontext- und Auswertungseinheit

Um die Genauigkeit der Inhaltsanalyse zu steigern, ist es notwendig Analyseeinheiten zu bestimmen (vgl. Mayring 2022:60). Dies geschieht auf den folgenden drei Ebenen:

● Kodiereinheit

Die Kodiereinheit definiert den kleinsten Bestandteil, der durch die Zuordnung zu den Kategorien ausgewertet werden darf (vgl. Mayring 2022:60). Für diese Arbeit sind dies die kleinsten bedeutungstragenden Elemente in den jeweiligen Interviews, wozu auch kürzeste Beiträge wie etwa Bejahungen von Fragen gehören sollen.

● Kontexteinheit

Die Kontexteinheit definiert, was das größte Textelement sein darf, das unter eine Kategorie fallen darf (vgl. Mayring 2022:60). Für diese Arbeit sollen das alle vollständigen Antworten auf eine Interviewfrage sein.

● Auswertungseinheit

Die Auswertungseinheit legt die Reihenfolge der Auswertung der einzelnen Textteile fest (vgl. Mayring 2022, 60). Für diese Arbeit sollen die drei Expert*innen-Interviews als Auswertungseinheiten festgelegt werden.

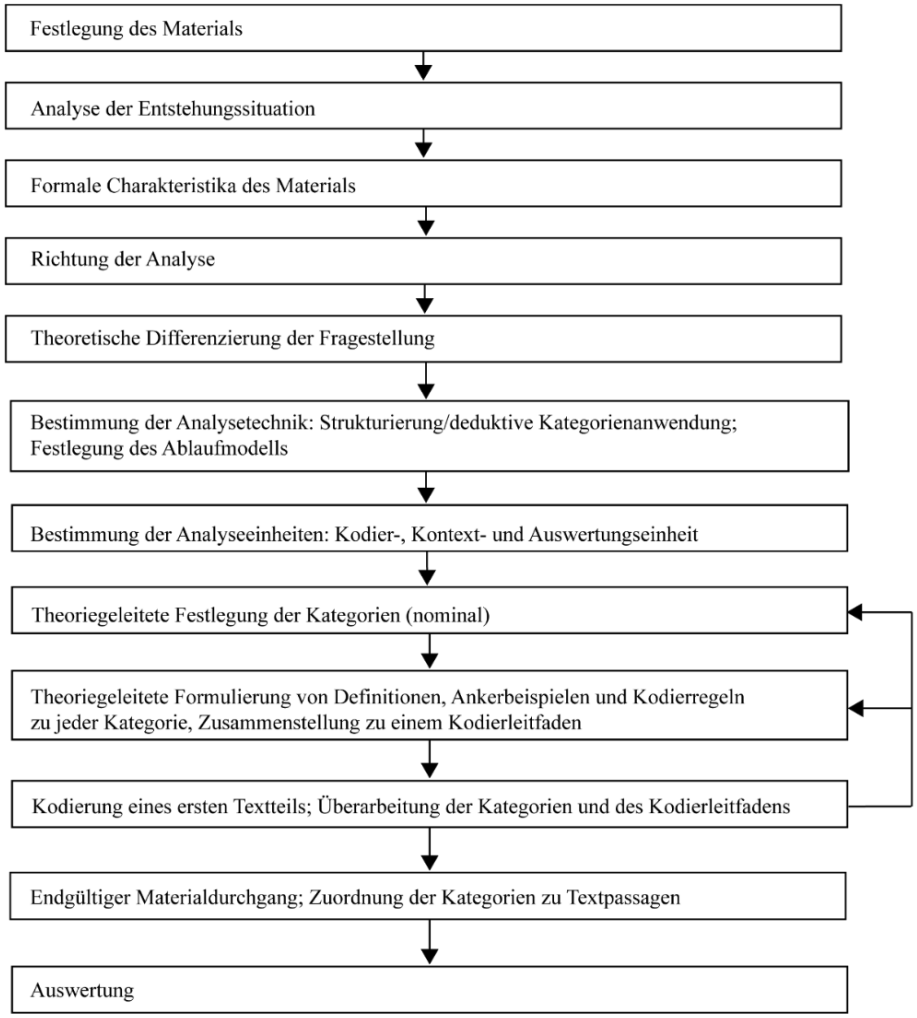

Abbildung 1: Ablaufmodell der Auswertung: eine Kombination des allgemeinen Ablaufmodells (vgl. Mayring 2022:61) und des Ablaufmodells der deduktiven Kategorienanwendung[10] (vgl. Mayring 2022:97).

[10] In Mayrings Modell für die Strukturierung/deduktive Kategorienanwendung findet vor der Auswertung noch ein Intercoder-Übereinstimmungstest statt. Dabei wird das Textmaterial von einer zweiten Person kodiert, um Reliabilität und Validität der Ergebnisse sicherzustellen. Dieser Schritt wird in dieser Arbeit ausgelassen, da es sich quasi um eine erste Pilotstudie zum Thema handelt und gegebenenfalls in weiteren Untersuchungen (vgl. Mayring 2022: 52f.) das Thema weiter erforscht wird (vgl. Mayring 2022:52f.).

h) Theoriegeleitete Festlegung der Kategorien für die Strukturierung/deduktive Kategorienanwendung

Bei der Analysemethode der Strukturierungen werden in einem deduktiven Prozess Kategorien auf das Material angewandt. Diese Kategorien müssen zunächst auf Basis der Theorie und der Fragestellung erarbeitet und definiert werden (vgl. Mayring 2022:67). Das Kategoriensystem selbst kann aus Hauptkategorien und Unterkategorien bestehen und nominal oder ordinal strukturiert sein (vgl. Mayring 2022:98). Für diese Arbeit wird ein nominales Kategoriensystem entwickelt, das in Haupt- und Unterkategorien strukturiert ist. Es befindet sich in Anhang 4. Ziel dieses Ansatzes ist die Herausarbeitung einer Struktur aus dem vorhandenen Textmaterial. Das im Vorfeld festgelegte Kategoriensystem dient dazu, diese Struktur sichtbar zu machen. Wenn bei der Analyse Textelemente unter eine bestimmte Kategorie fallen, werden diese aus dem Material systematisch herausgezogen (vgl. Mayring 2022:96).

i) Theoriegeleitete Formulierung von Definitionen, Ankerbeispielen und Kodierregeln zu jeder Kategorie, Zusammenstellung zu einem Kodierleitfaden

Der nächste Schritt besteht aus der Ableitung der Kategorien auf Basis der Theorie (vgl. Mayring 2022:97f.) und ihrer Definition. Für diese Arbeit ist das auf Basis der NISO-Prinzipien für gute Metadaten geschehen, die in Kapitel 3.3. vorgestellt wurden.

Anschließend wird das Material untersucht und für die Kodierung relevante Textelemente werden herausgefiltert und für den Kodierleitfaden als Ankerbeispiele verwendet. Sollten Textstellen nicht eindeutig einer Kategorie zugeordnet werden können, so muss festgelegt werden, welcher Kategorie diese Textstelle zugerechnet werden soll, und eine entsprechende Kodierregel aufgestellt werden. Dies erhöht die Genauigkeit des Kategoriensystems. Sollten vorherige Zuordnungen nun nicht mehr stimmen, müssen diese überarbeitet werden (vgl. Mayring:97f.). Dargestellt werden Definitionen, Ankerbeispiele und Kodierregeln in einem Kodierleitfaden (vgl. Mayring 2022:97.f.). Der Kodierleitfaden dieser Arbeit findet sich in Anhang 4.

j) Kodierung eines ersten Textteils und Überarbeitung der Kategorien und des Kodierleitfadens

Sobald Kategorien auf Basis des theoretischen Hintergrunds der Fragestellung definiert und Ankerbeispiele und Kodierregeln erstellt wurden, kann mit der Bearbeitung des Textmaterials begonnen werden: Als erstes wird nach Textteilen gesucht, die für die Fragestellung von Belang sind. Finden sich dabei Textstellen, die nicht kodiert werden kann, muss der Kodierleitfaden überarbeitet werden. (vgl. Mayring 2022:97).

k) Rückkoppelung zur theoriegeleiteten Festlegung der Kategorien für die Strukturierung/deduktive Kategorienanwendung

Als nächstes ist in Abbildung 1 ein Pfeil zurück zur theoriegeleiteten Festlegung der Kategorien für die Strukturierung/deduktive Kategorienanwendung zu sehen. Hier wird betrachtet, ob das Interviewmaterial Aspekte bietet, die noch nicht von den bisherigen Kategorien erfasst werden. Wenn dem so ist, werden zusätzliche Kategorien auf Basis der Literatur festgelegt und in den Kodierleitfaden aufgenommen. Daraufhin werden wieder Definitionen, Ankerbeispiele und Kodierregeln für jede weitere Kategorie, erstellt. Außerdem wird anhand des Interviewmaterials überprüft, ob nun die Kategorien alle relevanten Aspekte des Interviews abdecken. Falls nicht, gibt es eine weitere Schleife.

l) Endgültiger Materialdurchgang und Zuordnung der Kategorien zu Textpassagen

Wenn durch die neu formulierte Kodierregel frühere Kategorienzuordnungen falsch werden, so müssen auch diese revidiert werden. Auf diese Weise füllt sich der Kodierleitfaden immer mehr. Bei selbstevidenten und klar abgrenzbaren Kategorien können Ankerbeispiele und Kodierregeln auch ausgelassen werden. Wenn sich die Kodierarbeit stabilisiert (keine neuen Kodierregeln werden nötig), ist es Zeit für den allgemeinen Revisionsdurchgang (beschrieben durch den Rückkoppelungspfeil im Ablaufmodell). Die bisherige Kodierarbeit war in diesem Fall die Pilotstudie (vgl. Mayring 2022:98).

m) Auswertung

Abschließend findet die Auswertung des Textmaterials statt: dafür werden die deduktiv gebildeten Kategorien den relevanten Textstellen zugeordnet[11] (vgl. Mayring 2022:98). Dieses Kapitel hat die für diese Arbeit eingesetzten Methoden, die Erhebungsmethode der Expert*innen-Interviews und die Methode der qualitativen Inhaltsanalyse, die der Auswertung dient, vorgestellt. Für beide Methoden wurden die einzelnen notwendigen Schritte dargelegt, die zunächst für die Erstellung des Interviewleitfadens und die Auswahl der Expert*innen für diese Arbeit und dann die Auswertung der Interviews notwendig sind.

[11] Die Textstellen, die schließlich relevant für die Auswertung der Expert*innen-Interviews waren, sind in den Interviews im Anhang grau markiert.

5. Auswertung der Expert*innen-Interviews

In diesem Kapitel findet die Darstellung der durch die Expert*innen-Interviews gewonnen Informationen statt. Dafür wurden die Expert*innen in teilstrukturierten Leitfaden-Interviews zu unterschiedlichen Aspekten der Arbeit mit Metadaten befragt, die hier als Basis für die Strukturierung verwendet werden. Der Interviewleitfaden findet sich in Anhang 1. Die Interviews werden mittels der qualitativen Inhaltsanalyse nach Mayring (1983/2022) ausgewertet. Für die Auswertung wurden die Kategorien des Kodierleitfadens (s. Anhang 4) als roter Faden verwendet.

5.1. Verwendung von Community Standards

In den folgenden Abschnitten soll anhand der Expert*innen-Interviews betrachtet werden, ob und wie sich die Verwendung von Community Standards und Application Profiles auf die Arbeit mit Metadaten auswirkt.

5.1.1. Auswahl des Schemas mit Bezug auf die in der Community verwendeten Schemata

Der Einsatz von in der jeweiligen Community verwendeten Standards dient dazu, die Metadaten eines digitalen Repositoriums anderen Institutionen zugänglich zu machen und auch Interoperabilität herzustellen.

In PHAIDRA wurde bei der Arbeit mit den Metadaten von Beginn an darauf geachtet, auch die in der Community verwendeten Schemata einzusetzen, wie Blumesberger, bemerkt:

> „[W]ir haben von Anfang an einige Metadaten-Schemata drinnen gehabt in PHAIDRA, Dublin Core zum Beispiel, und das LOM-Schema – Learning Object Metadata. Das hatten wir deshalb drinnen, weil wir auch die Lehre sehr dabeihaben wollten. Und das hat sich jetzt als sehr günstig erwiesen, weil inzwischen gibt es, also eh schon länger, Projekte zu Open Educational Resources, die genau dieses Metadatenschema benötigen, um eben diese Objekte gut beschreiben zu können“ (Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 2.2.).

Die Entscheidung für das LOM-Schema war Resultat der kooperativen Entwicklung des Repositoriums PHAIDRA durch die Universitätsbibliothek, den Zentralen Informatikdienst sowie das Center for Teaching and Learning der Universität Wien, wobei letzteres

die Perspektive der Lehre (LOM-Schema) mit in die Entwicklung einbrachte (vgl. Ganguly, Raman, Expert*innen-Interview, Wien, 20.12.2023, s. Anhang 2.1.). Der Einsatz der in der Community verbreiteten Metadaten-Schemata, wie etwa dem LOM-Schema, bringt auch heute noch Vorteile, wenn es etwa um Kooperationen im Bereich der offenen Bildungsressourcen geht. Damit könnte man das Thema als abgeschlossen betrachten.

Bei der Entwicklung des Repositoriums PHAIDRA und der dort verwendeten Metadatenkategorien zur Beschreibung der digitalen Objekte durch die Universitätsbibliothek, den Zentralen Informatikdienst sowie das Center for Teaching and Learning der Universität Wien, wurde außerdem versucht den unterschiedlichen Bedarfen und spezifischen Wünschen der verschiedenen Nutzer*innen-Gruppen gerecht zu werden. Dafür wurde eine breite Skala an Metadatenkategorien verwendet, die die unterschiedlichen Arten von digitalen Objekten beschreiben konnte. Dadurch ist es schließlich „zu einem sehr großen (…), nicht einem Standard entsprechenden Metadatenkonglomerat gekommen, das wir dann UW-Metadaten genannt haben“ (Ganguly, Raman, Expert*innen-Interview, Wien, 20.12.2023, s. Anhang 2.1.). Ganguly nennt dies auch „eine Zweckgemeinschaft von verschiedenen Standards“ (Ganguly, Raman, Expert*innen-Interview, Wien, 20.12.2023, s. Anhang 2.1.). In dem Format UW-Metadaten waren die von den unterschiedlichen Nutzer*innen-Gruppen benötigten Metadatenkategorien enthalten und es war dadurch möglich, die vielfältigen Forschungsarbeiten und Lehrperspektiven an der Universität Wien zu umfassen. Gleichzeitig war dies der Ausgangspunkt der Arbeit mit den unterschiedlichen Metadaten-Schemata der Community. Laut Ganguly wurden vom Format UWMETADATEN ausgehend „verschiedene Standards herausdestilliert […] Das heißt, wir haben immer gewisse Ausschnitte. Was zwar in der Handhabung sehr schwierig ist, was uns aber jetzt in der Änderung der verschiedensten Szenarien, wo wir das Ganze einsetzen wollen, beziehungsweise wo wir überall auch die Metadaten hinschicken möchten, uns etwas leichter tun“ (Ganguly, Raman, Expert*innen-Interview, Wien, 20.12.2023, s. Anhang 2.1.). Dies bedeutet, dass aus dem Format UWMETADATEN auf diese Standards gemappt wurde. Auch hier zeigt sich, einzelne Standards der Community erfüllen nicht alle Bedarfe, doch ist man in der Lage mit mehreren Community-Standards zu arbeiten, sind sie auch nützlich für Kooperationen.

Die Verwendung von UWMETADATA ersetzt dabei nicht den Einsatz der Community Standards. Vielmehr ergänzt sie die Community Standards, denn „meistens will man ein Dublin Core sehen, oder ein bestimmtes Metadaten-Schema wie OAI-PMH, OpenAIRE oder DC Terms und auf die müssen wir immer exportieren." (Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.). Sehen bedeutet in diesem Zusammenhang, dass sie von Maschinen abgerufen werden und damit Interoperabilität ermöglichen. Von Beginn an wurden deshalb in PHAIDRA Standards der Community verwendet, etwa, wie bereits erwähnt das Learning Object Metadata-Schema und Dublin Core. Zusätzlich wurde auch DataCite eingesetzt (vgl. Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 2.2.).

Zwischen dem eigens entwickelten Metadatenschema für die zahlreichen unterschiedlichen Disziplinen der Universität Wien und den verwendeten Community Standards hält sich jedoch ein Spannungsfeld: Denn auch wenn Community Standards wie Dublin Core sehr nützlich sind, so erfüllen sie nicht alle Bedürfnisse, die an die Beschreibung von digitalen Objekten in PHAIDRA und anderen digitalen Repositorien gestellt werden. „Dublin Core war sehr erfolgreich, aber auch nur deswegen, weil es extrem vereinfacht ist. Und das ist nicht ein Format, in dem man wirklich die Metadaten speichern kann, weil da kann man nicht / da kann man nur sagen „creator" oder „contributor" und da sagt der Bibliothekar, das geht nicht so, da brauche ich schon ein bisschen mehr." (Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.). Dies bedeutet, dass nur grundlegende Informationen von den Community Standards, wie etwa Dublin Core, erfasst und damit auch ausgetauscht werden können. Viele andere Informationen, die für die Verwendung und das Verstehen der digitalen Objekte in PHAIDRA notwendig sind, können darin nicht festgehalten werden. Daher sind zusätzliche Metadatenfelder notwendig, die auch den Bedarfen der Forscher*innen gerecht werden und im UWMETADATEN-Schema festgehalten wurden (vgl. Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.).

Wenn digitale Objekte unterschiedlicher Fachbereiche in einem digitalen Repositorium enthalten sein sollen, muss dieses flexibel bleiben, um den unterschiedlichen wissenschaftlichen Disziplinen und ihren Ansprüchen an die Beschreibung ihrer digitalen Objekte auf geeignete Art entsprechen zu können. Dies unterstreicht auch Hudak, wenn

er sagt, dass Repositorien wie PHAIDRA sich nicht spezialisieren (vgl. Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.). Es ist damit schwierig ein verbreitetes Metadatenschema zu finden, das sowohl den Bedarfen an eine detaillierte Beschreibung der zahlreichen und vielfältigen Disziplinen an einer großen Universität als auch dem Wunsch nach Interoperabilität mit anderen Repositorien gerecht wird. Der Einsatz von Community Standards allein ist daher keine allgemeingültige Lösung, wenn es darum geht, digitale Objekte zu beschreiben, auch weil es immer viele unterschiedliche Standards geben wird (vgl. Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.). Dies bedeutet auch, dass Mappings immer wieder notwendig sein werden.

Betrachtet man die Frage, mit welchen anderen Repositorien man interoperabel sein möchte, aus der Perspektive der Vielfalt unterschiedlichster Institutionen und ihrer jeweiligen Bedürfnisse, deutet dies auch auf die begrenzte Möglichkeit der Verwendung von Standards. Hudak bemerkt dazu:“[E]s gibt diesen Spruch, die Standards sind wie Zahnbürsten, dass jeder denkt, es ist eine super Idee, einen Standard zu verwenden, aber niemand will den Standard von den anderen verwenden. […] Ich kann mir nicht vorstellen, dass es irgendeine Entität gibt, so auch die Library of Congress oder wer auch immer, der sagt, das ist jetzt der Standard und alle werden diesen Standard verwenden. Das kommt nicht. Und deswegen ist dann immer etwas nicht interoperabel und da hätte ich keine bessere Idee als das RDF und das Linked Data“ (Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.). Am Ende dieses Zitats wird deutlich, dass Linked Data oder RDF eine Möglichkeit sein könnten, um einerseits den Bedarfen unterschiedlicher Institutionen an ihre Metadaten, aber gleichzeitig auch dem Wunsch nach möglichst umfassender Interoperabilität gerecht zu werden.

Hudak (vgl. Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.) stellt fest, dass es bei der Verwendung von RDF zu denselben Problemen kommen kann, wie bei der Arbeit mit Metadaten-Standards wie MODS, UWMETADATA oder LOM, die in PHAIDRA eingesetzt werden. Diese Probleme können in RDF etwa durch die Begrenzung der Metadatenfelder auf einen bestimmten Bereich entstehen. Wie bei der Arbeit mit einem XML-Metadaten-Schema, fehle bei einer auf die eigene Thematik beschränkten Benutzung von RDF hin und wieder ein Feld, es gäbe unterschiedliche

Interpretationen. Bei XML-Metadaten-Schemata müssten die Standards und Felder wieder gemappt werden (vgl. Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.). Dies müsse bei der Arbeit mit RDF jedoch nicht unbedingt geschehen, da RDF es ermöglicht, Elemente unterschiedlicher Standards zu kombinieren:

> „man nimmt sich die Metadatenfelder, aus den verschiedenen Standards. Und zwar ist es praktisch und flexibel, weil man (…) kann die einfach nehmen. Aber auch die Idee dahinter ist diese Linked Data-Idee, dass ich nicht wieder einen neuen Namen für den Titel hab, sondern ich nehme den, den die meisten Leute kennen. Das heißt, ich sag „Dublin Core Title" und das sollte dann schon sehr vielen anderen Maschinen etwas sagen" (Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.)

Dies bedeutet, dass aus den unterschiedlichen und am meisten verbreiteten Standards die Metadatenfelder ausgewählt werden und in die eigene Darstellungsweise eingearbeitet werden können. Damit wird die Verwendung von weit verbreiteten Standards, die auch andere Repositorien über ihre Harvesting-Protokolle verstehen können, sichergestellt, was die Interoperabilität fördert. Gleichzeitig finden durch die große Auswahl an möglichen Standards und damit möglichen Metadatenfeldern auch die Ansprüche der Nutzer*innen des jeweiligen Repositoriums Berücksichtigung. Im Fall von PHAIDRA kann so den Wünschen und Bedarfen einer sehr diversen Nutzer*innengemeinschaft entsprochen werden.

Einen kleinen Einblick, welche Vielfalt an Metadatenfeldern dies umfasst, gibt folgendes Zitat von Hudak: „Also wir haben momentan, glaube ich, mehr als 50 sogar Ontologien drinnen. Schema, BIBFRAME[12], ich weiß nicht, EBUCore, für diese Dinge wie MIME type oder so, und (…) RDA[13] auch, für exotischere Sachen, die haben sehr viele Prädikate" (Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.). Diese Vielfalt ermöglicht eine genaue Beschreibung der digitalen Objekte durch Metadaten in PHAIDRA, wie sie von den unterschiedlichen Fakultäten und Instituten der Universität Wien benötigt wird. Hudak formuliert es knapp so: „jedes Feld haben wir von irgendwoher genommen, so wie wir es gebraucht haben" (Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.). Beispielhaft erwähnt Hudak (vgl. Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.). die

[12] BIBFRAME steht für: Bibliographic Framework Initiative.
[13] RDA steht für: Resource Description and Access.

Erweiterung der Metadatenfelder zu den durch das digitale Objekt repräsentierten Objekten, etwa wenn physische Objekte fotografiert werden, und dann durch die Metadaten beschrieben werden sollen. Dabei kann es um Metadatenkategorien wie die physische Größe des abgebildeten Objekts gehen. Auch bei der Arbeit mit digitalisierten DVDs ist es sehr nützlich, wichtige Informationen – etwa die Zielpublikum eines Film – durch zusätzliche Metadatenfelder festhalten zu können (vgl. Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.).

5.1.2. Verwendung von Application Profiles

Wenn es darum geht, Community Standards für die Beschreibung digitaler Objekte in digitalen Repositorien zu verwenden, können diese auch immer entsprechend der eigenen Bedürfnisse angepasst werden. Wie selbstverständlich das in der Arbeit mit den Metadatenformaten in PHAIDRA ist, zeigt folgendes Zitat von Hudak: „Was die internen Metadatenformate betrifft, das LOM-Metadaten-Schema wurde angepasst/erweitert und das JSON-LD-Schema wird oft nach Art des Bestands anders verwendet (eine andere Untermenge von Feldern, bestimmte kontrollierte Vokabulare, etc.). Man kann das auch Application Profiles nennen.“ (Hudak, Rastislav, Expert*innen-Interview per Mail, Wien, 15.03.2024, s. Anhang 3.2.). Auch Blumesberger bestätigt den Einsatz von Application Profiles bei der Arbeit mit PHAIDRA (vgl. Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 3.1.). Diese Anpassung an die Bedarfe des jeweiligen Repositoriums und der eigenen Organisation wird sich in der Arbeit mit dem RDF-Format weiter fortsetzen, wie das letzte Hudak-Zitat zeigt. Die Anpassung von Metadaten-Schemata an die eigenen Bedarfe scheint daher immer eine Möglichkeit und auch Notwendigkeit zu sein. Allerdings könnte der Einsatz von Application Profiles Mappings notwendig machen, zwischen den im Ausgangspunkt gleichen Metadaten-Schemata, die von zwei Organisationen jeweils unterschiedlich interpretiert wurden (vgl. Ganguly, Raman, Expert*innen-Interview, Wien, 20.12.2023, s. Anhang 2.1.).

5.2. Unterstützung von Interoperabilität

In den folgenden Abschnitten soll auf Basis der Expert*innen-Interviews dargestellt werden, wie sich unterschiedliche Aspekte, die Interoperabilität fördern sollen, auf die Arbeit mit Metadaten auswirken. Dazu gehören die detaillierte Vergabe von Metadaten, die Zusammenarbeit mit Aggregatoren, die Speicherung von Metadaten in Repositorien und ihre generelle Dokumentation und Zugänglichkeit, die Erstellung von Mappings selbst, und die Verwendung von Harvesting-Protokollen.

5.2.1. Vergabe von detaillierten Metadaten als Grundlage für Interoperabilität

Metadaten dienen dazu, Objekte möglichst so zu beschreiben, dass ihre Deutbarkeit auch außerhalb der Institution, an der sie entstanden sind, sichergestellt ist.

Bei der Entwicklung des Repositoriums PHAIDRA an der Universität Wien spielte die detaillierte Vergabe von Metadaten von Beginn an eine wichtige Rolle. Wie zentral dieser Aspekt war, zeigt folgendes Zitat von Susanne Blumesberger:

> „Also unser Repositorium wurde ja 2008 sozusagen in die Welt gesetzt. Seitdem gibt es das Repositorium. Und die Arbeit an den Metadaten, also mal zu überlegen, welche Metadaten man überhaupt benötigt in einem Repositorium, das war ein ganz ein wichtiger Punkt und hat eigentlich das erste Jahr richtig ausgefüllt. Also wir haben 2007 begonnen. Da startete das Projekt, so hieß es damals: Projekt Digitale Langzeitarchivierung. Und wir haben wirklich ein Jahr lang versucht herauszufinden, welche Metadaten von welchen Personen gebraucht werden.“ (Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 2.2.).

Wichtig ist dabei festzuhalten, dass eine große Universität wie die Universität Wien eine Vielzahl an Disziplinen und damit auch Anforderungen an Metadaten umfasst. Um dieser Vielfalt der Fachbereiche und Ansprüche gerecht werden zu können, hat das PHAIDRA-Team von Beginn an unterschiedliche Stakeholder der Universität miteinbezogen, wie Susanne Blumesberger berichtet:

> „wir sind dann wirklich in der Uni Wien herumgegangen, haben uns Pilotpartner gesucht. Aus unterschiedlichen Bereichen, zum Beispiel aus der Translationswissenschaft, aus der Bibliothek, aber auch aus der Physik, anderen Bereichen, und haben einfach geschaut: Was brauchen die Wissenschaftler? Was brauchen die Forschenden, um eben ihre Objekte gut beschreiben zu können? (vgl. Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 2.2.).

Zentral war dabei der Blick auf die Bedarfe der Forschenden, denen es ermöglicht werden sollte, die Objekte so detailliert zu beschreiben, dass sie den Anforderungen der einzelnen Disziplinen gerecht werden. Gleichzeitig mussten auch Ziel und Zweck eines digitalen Repositoriums auf Universitätsebene berücksichtigt werden. Die Schwierigkeiten, die bewältigt werden mussten und müssen, um dieser Diversität gerecht zu werden, werden auch hier deutlich:

> „die Uni Wien ist riesig und sehr, sehr heterogen. Das heißt, wir haben unglaublich unterschiedliche Daten und die alle unter einen Hut zu bringen und zu sagen, ich kann die jetzt einfach in einem Repositorium so beschreiben, dass man wirklich das Gefühl hat, man hat die Daten gut beschrieben und das kommt eigentlich dem zugute, was man wirklich auch möchte. Das heißt, man trifft wirklich genau den Punkt, ist gar nicht so einfach, und das war für uns eine ganz wichtige Phase natürlich.“ (vgl. Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 2.2.).

Wie unterschiedlich die Anforderungen der verschiedenen Institute sind, wird sichtbar, wenn man etwa die Anforderungen betrachtet, die das Institut für Kunstgeschichte an die Metadaten in PHAIDRA stellte: Es wünschte sich Stempel durch Metadaten beschreiben zu können, die auf bestimmten digitalisierten Glasplatten-Dias vorhanden sindund eine zentrale Information für die Forschenden darstellen (vgl. Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 2.2.).

Ein weiteres Beispiel dafür, wie detailliert Metadaten in PHAIDRA vergeben werden können, sind digitale Objekte, auf denen etwa ein Gemälde abgebildet ist: Diese können in PHAIDRA auf Ebene des digitalen Objekts beschrieben werden und darüber hinaus auch Angaben zum digitalisierten Objekt, also zum Gemälde selbst, etwa zur Entstehungszeit oder zum jeweiligen Maler des Gemäldes, in den Metadaten festhalten (vgl. Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 2.2.).

Um diesen Anforderungen gerecht zu werden, wurde das Metadatenformat UWMETADATA verwendet, da es in der Lage war, den vielfältigen Bedarfen der Forschenden, Lehrenden und Studierenden an der Universität Wien zu entsprechen und die digitalen Objekte möglichst detailliert zu beschreiben.

Doch die detaillierte Vergabe von Metadaten in PHAIDRA beschränkt sich nicht auf die genaue Beschreibung von digitalen Objekten und den Objekten, die auf ihnen

abgebildet sind. Zusätzlich werden auch, wenn es notwendig ist, die Pflicht-Metadatenfelder angepasst, um eine möglichst gute Beschreibung der digitalen Objekte zu gewährleisten und die Wiederauffindbarkeit zu verbessern (vgl. Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 2.2.). Außerdem wurden für unterschiedliche Arten von Projekten auch eigene Masken mit entsprechenden Metadatenfeldern vom Zentralen Informatikdienst erstellt, die auch in Zukunft an die jeweiligen Bedarfe angepasst werden können (vgl. Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 2.2.) und damit die Beschreibung der digitalen Objekte und ihre Weiter- und Wiederverwendung sicherstellen.

Da PHAIDRA ein Repositorium für die gesamte Universität Wien mit all ihren Disziplinen ist, hatte die erste Metadaten-Eingabemaske etwa 150 Felder, die man alle beim Upload eines digitalen Objekts durchgehen musste. Dadurch konnten die digitalen Objekte genau mit Metadaten beschrieben werden. Jedoch war dies dann zu umfassend, und diese Eingabemaske wurde überarbeitet. Nun gibt es in der Eingabemaske nur noch etwa 10 Metadatenfelder, die für die Beschreibung der digitalen Objekte verpflichtend ausgefüllt werden müssen (vgl. Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.). Zusätzlich können aktuell auch weitere Metadaten vergeben werden, indem Bereiche ausgeklickt werden, die ergänzende Informationen zu den digitalen Objekten umfassen können (vgl. Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3).

Im Bereich der Vergabe von Metadaten gibt es auch ein Spannungsfeld zwischen den Wünschen der Forschenden und den Möglichkeiten und Aufgaben eines digitalen Repositoriums. Während Forschende ihre digitalen Objekte teilweise geradezu wissenschaftlich beschreiben wollen, kann es aus Sicht der Repositorienbetreiber*innen notwendig sein, pragmatisch zu bleiben und die Auffindbarkeit der digitalen Objekte ins Zentrum zu stellen. Zusätzliche für die Forschenden relevante Informationen können dann, auch aus Gründen des Urheberrechtsschutzes, in einem eigenen PDF zur Verfügung gestellt und ebenfalls in PHAIDRA als digitales Objekt hochgeladen werden (vgl. Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.).

Die detaillierte Vergabe von Metadaten ist auch für das Institut für Kultur- und Sozialanthropologie gegenwärtig ein sehr wichtiges Thema. Dieses würde gerne in den

Metadaten zusätzliche Informationen zu den unterschiedlichen Ebenen eines digitalen Objekts abbilden können. Blumesberger nennt als Beispiel einen Webstuhl, der auf einem Glasplattendia abgebildet ist. Nun sollen das digitale Objekt selbst, das Glasplattendia, das darauf abgebildet wird, und auch der auf dem Glasplattendia abgebildete Webstuhl in den Metadaten in PHAIDRA beschrieben werden (vgl. Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 2.2.). Dies wäre eine sehr komplexe Beschreibung und zum Teil im Metadatenformat UWMETADATA möglich, jedoch wären dafür einige zusätzliche Felder notwendig und daher fiel die Entscheidung, diese Art von Beschreibung in die Entwicklung des neuen Metadatenformats in RDF zu verlagern und dieses so möglichst zügig weiterzuentwickeln, wie Hudak bemerkt (vgl. Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.).

Ein Ansatz, der betrachtet werden muss, um in Zukunft eine gute detaillierte Beschreibung der digitalen Objekte für die unterschiedlichen Fachbereich zu ermöglichen, setzt bei der Gestaltung der Eingabemaske beim Upload eines digitalen Objekts an, wie Hudak beschreibt:

> „Ansonsten arbeiten wir eher an der Infrastruktur als an dem Metadatenformat an sich und daran, dass man die Metadaten, beziehungsweise die Eingabemasken sich selbst gestalten kann, per quasi Instanz. Man kann es sich schon jetzt selber gestalten mit Templates. Aber oft ist es jetzt der Fall, dass ein Institut schon sagt, okay, wir brauchen eine generelle Eingabemaske, aber wir brauchen auch eine Eingabemaske für offene Bildungsressourcen, für Forschungsdaten, das sollte was anderes sein. Und dass das irgendwie so geht, dass wir nicht in jeder Instanz alles extra implementieren müssen, sondern dass man sich selbst einstellen kann, das sind jetzt die nächsten Schritte, also die sind nicht, die haben jetzt nicht unmittelbar mit dem Metadatenformat was zu tun.“ (Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.).

Die Eingabemaske soll es in Zukunft zulassen, diejenigen Metadatenfelder auszuwählen, die für das digitale Objekt relevant sind. Dadurch könnte den Bedarfen des jeweiligen Uploads rund um das jeweilige digitale Objekt entsprochen werden. Bisher ist es zwar schon möglich, Templates beim Upload zu verwenden, doch reicht dies nicht, um den unterschiedlichen Anforderungen gerecht zu werden, denen die unterschiedlichen Fachbereiche der Universität begegnen, wenn sie mit verschiedenartigen digitalen Objekten arbeiten (Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.).

Die Konsequenz, mit der an der Universität eine detaillierte Beschreibung der digitalen Objekte durch Metadaten ermöglicht und umgesetzt wird, ist sehr deutlich geworden. Sie reicht vom Mit-Einbeziehen unterschiedlicher Institute als Pilotpartner bei der Entwicklung von PHAIDRA und dem UWMETADATA-Format über die Zusammenarbeit mit einzelnen Instituten, um ihren Bedarfen gerecht zu werden, bis zur Anpassungsmöglichkeit von Eingabemasken.

Doch Ganguly sieht trotz der bisherigen Arbeit noch die Möglichkeit für weitere wichtige Entwicklungen, um die digitalen Objekte zugänglich und in anderen Kontexten oder auch Disziplinen interpretierbar zu machen und stellt fest: „Das heißt, wir müssen wahrscheinlich, oder das ist der Trend, den ich sehe, mehr Flexibilität auch in der Beschreibung der Metadaten oder der Objekte halt zu haben, im Sinne, die sich dann in den Metadaten niederschlägt, um die Findability zu erhöhen“ (Ganguly, Raman, Expert*innen-Interview, Wien, 20.12.2023, s. Anhang 2.1.). Er beschreibt auch das Problem genauer, sowie einen Lösungsansatz:

> „Das ist einerseits, dass die Objekte mehrfach beschrieben werden. Das heißt, von unterschiedlichen Communities unterschiedlich beschrieben werden. Das ist insofern auch interessant für Objekte, die dann in unterschiedlichen Kontexten auch verwendet werden. Zum Beispiel gibt es ja gerade in der Lebenswissenschaft sehr viel interdisziplinäre Forschung. Was jetzt bedeutet, dass zum Beispiel Forschungsergebnisse aus der Chemie auch für die Biologie interessant sind, umgekehrt auch wieder für die Biologen, die Chemie. Und dann diese Forschungsbereiche, die sowieso beides inkludieren müssen, wie zum Beispiel die Pharmazie, die einerseits von der Biologie Daten hat und andererseits auch von der Chemie die Daten hat. Die Schwierigkeit in dem Bereich ist es, dass Chemiker die Dinge anders beschreiben als Biologen und umgekehrt. Und wenn jetzt diese Objekte, die relevant sind, aus mehreren Ebenen beschrieben werden, dann wären sie natürlich für mehrere Communities leichter zu finden, weil einfach gewisse Terminologien verwendet werden“ (Ganguly, Raman, Expert*innen-Interview, Wien, 20.12.2023, s. Anhang 2.1.)

Eine mehrfache Beschreibung der Objekte aus den Blickwinkeln der unterschiedlichen Disziplinen, um Interdisziplinarität zu fördern, würde sich auch auf die Auswahl der Metadatenfelder während des Uploads sowie die Eingaben in die Metadatenfelder auswirken, ist jedoch immer mit zusätzlichem Aufwand verbunden (vgl. Ganguly, Raman, Expert*innen-Interview, Wien, 20.12.2023, s. Anhang 2.1.). Die Beschreibung der digitalen Objekte aus mehreren Perspektiven erleichtern würde der Einsatz von KI, wie Ganguly bemerkt:

> „Ich denke, dass in Zukunft, und das wird jetzt der spannende Bereich, auch für die Repositorien-Welt und für die Datenwelt sein, hier den Einsatz von KI zu forcieren, nämlich das automatisierte Beschreiben von diesen Objekten, und das aus unterschiedlichen Perspektiven zu sehen. Also das sehe ich auch als Zukunftstrend, noch nicht aktuell, weil momentan ist es ein großer Hype und man muss schauen, was da übrigbleibt und nicht bei jedem Hype halt dann mitspringen, sondern wirklich schauen, wo ist dann auch die Qualität dahinter?“ (vgl. Ganguly, Raman, Expert*innen-Interview, Wien, 20.12.2023, s. Anhang 2.1.).

5.2.2. Zusammenarbeit zwischen Organisationen und Aggregatoren für Interoperabilität

Aggregatoren spielen eine wichtige Rolle, wenn es um die Auffindbarkeit von digitalen Objekten in digitalen Repositorien geht. Bei der Entwicklung von PHAIDRA stellte die Zusammenarbeit mit Aggregatoren von Beginn an einen wichtigen Aspekt dar, wie Blumesberger berichtet:

> „es war von Anfang an mitgedacht, dass wir ja, also einen Austausch haben wollen mit anderen Systemen. Und da waren wir recht früh auch mit OpenAIRE zum Beispiel in Kontakt. Also ich glaub seit 2009, dieses EU-Projekt, die auch harvesten, ja, die sich auch Daten holen und Europeana. Das war auch / vor allem Europeana Libraries, das war ein großes Projekt bei uns. Das uns sehr geholfen hat, weil da sehr viel digitalisiert werden konnte und andererseits natürlich auch genau dieser Austausch angeregt werden konnte. Ja, also das, das war von Anfang an ein großer Punkt. Das haben wir immer gesagt, dass das wichtig ist, ja“ (Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 2.2.)

Um die Metadaten aus PHAIDRA mit anderen Systemen auszutauschen, kooperiert PHAIDRA etwa mit den Aggregatoren OpenAIRE und Europeana. Dabei holen sich OpenAIRE und Europeana die Metadaten aus PHAIDRA über ein Harvesting-Protokoll. Dafür müssen die Metadaten jedoch in einem bestimmten Format vorhanden sein. Auf diese Art nehmen sich Aggregatoren zwar nur die grundlegenden Daten, wie etwa „Titel, eine Beschreibung, Autor oder Datum. […] und die Lizenz natürlich“ (Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 2.2.), die Kooperation mit Aggregatoren ist für PHAIDRA und andere Repositorien jedoch selbstverständlich (vgl. Ganguly, Raman, Expert*innen-Interview, Wien, 20.12.2023, s. Anhang 2.1.). Die Metadaten, die dabei weitergegeben werden, unterstützen die Auffindbarkeit der digitalen Objekte. Wenn es jedoch darum geht, die Objekte detaillierter zu beschreiben, werden die dazu notwendigen Metadaten nicht weitergegeben und bleiben in PHAIDRA. Allerdings bieten sie dort für die Nutzer*innen einen Mehrwert, da sie die digitalen Objekte genau erfassen (vgl. Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024,

s. Anhang 2.2.). Wenn Metadaten an Aggregatoren weitergegeben werden, kann daher nicht sichergestellt werden, dass diese eine detaillierte Beschreibung der Objekte erhalten, wie sie in PHAIDRA möglich und teilweise für wissenschaftliches Arbeiten auch relevant ist. Die Verbreitung der Metadaten und damit Informationen zu digitalen Objekten, schließt daher ihre detaillierte Beschreibung nicht unbedingt mit ein. Die Kooperation mit Aggregatoren ist jedoch ein wichtiger Bereich bei der Arbeit mit PHAIDRA und damit die Aggregatoren die Metadaten abholen, müssen diese in einem bestimmten Format vorhanden sein.

5.2.3. Interoperabilität unterschiedlicher Metadaten-Schemata durch ihre Speicherung in Repositorien und generell öffentliche Dokumentation und Zugänglichmachung

Um Metadaten-Schemata und damit in ihnen gespeicherte Metadaten weiter- und wiederverwenden zu können, ist es nützlich, erstere intern zu dokumentieren und ihre Dokumentation auch öffentlich zugänglich zu machen. Dies kann in dafür vorgesehenen Repositorien geschehen. Jedoch auch auf andere Art und Weise, wie etwa die Dokumentation auf frei zugänglichen Webseiten.
Wenn es um die Zugänglichkeit der Metadaten-Schemata in PHAIDRA geht, sind einige der dort verwendeten Schemata öffentlich zugänglich und gut dokumentiert, etwa Dublin Core. Allerdings nicht in einem dafür vorhergesehenen Repositorium, sondern auf einer eigenen Webseite.

Das UWMETADATA-Schema selbst ist zwar intern gut dokumentiert (vgl. Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 2.2.), jedoch nur über die digitalen Objekte, deren Metadaten in diesem Format gespeichert wurden, für die Öffentlichkeit zugänglich. Eine ausführliche Dokumentation für die Öffentlichkeit wurde nicht zur Verfügung gestellt. Ohne weitere Informationen könnte ein externes System daher das Metadatenformat UWMETADATA nur in einem begrenzten Ausmaß verstehen (vgl. Hudak, Rastislav, Expert*innen-Interview per Mail, Wien, 15.03.2024, s. Anhang 3.2.). UWMETADATA blieb daher immer ein Metadatenformat, das vor allem an der Universität Wien eingesetzt wurde. Dadurch wurden auch Mappings in Community Standards und für die Kooperation mit Aggregatoren und Projektpartnern generell notwendig, wie Kapitel 5.2.1. und 5.2.2. zeigten. Andere Metadaten-Schemata, wie etwa

MODS der Library of Congress haben Verbreitung gefunden und werden von anderen Repositorien eingesetzt (vgl. Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.).

Die komplexe Situation, die durch die unterschiedlichen Fachbereiche und ihre Anforderungen an Metadaten an der Universität Wien entsteht, machte die Entwicklung eines eigenen Schemas notwendig. Dieses wurde nicht eingesetzt und verbreitet, um Interoperabilität mit anderen Institutionen zu ermöglichen, sondern vor allem, um die eigenen digitalen Objekte möglichst gut zu beschreiben. Es ist aber intern gut dokumentiert (vgl. Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 2.2.).

Durch das neue Metadatenschema in PHAIDRA, JSON-LD[14] im RDF-Format, konzentriert man sich jedoch auf die Verwendung von bereits bestehenden Metadatenformaten, die gut dokumentiert sind, und deren Elemente entsprechend den Notwendigkeiten in PHAIDRA eingesetzt werden können (vgl. Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.). Dadurch ergibt sich, dass in Zukunft sowohl die globale Interoperabilität gesteigert als auch die detaillierte Beschreibung von digitalen Objekten ermöglicht wird. Denn auf die global gültige Dokumentation und Definition der verwendeten Standards, die nun meist Ontologien sind, wird mittels eines Links in den Metadaten im JSON-LD-Format verwiesen. Dort findet man dann alle Informationen dokumentiert (vgl. Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.). Darüber hinaus sind die verwendeten Metadaten-Standards auch auf einer Wiki-Seite dokumentiert (vgl. Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.).

5.2.4. Interoperabilität unterschiedlicher Metadaten-Schemata durch Mappings

Der Einsatz unterschiedlicher Metadaten-Schemata und Mappings, um Interoperabilität zu ermöglichen, ist ein wichtiger Aspekt in der Arbeit mit PHAIDRA. Wie bereits in Kapitel 5.1.1 von Ganguly (vgl. Ganguly, Raman, Expert*innen-Interview, Wien, 20.12.2023, s. Anhang 2.1.) erwähnt, wurden aus UWMETADATA unterschiedliche

[14] JSON-LD ist das Akronym für Java Script Object Notation for Linked Data

Standards extrahiert. Dafür waren auch Mappings notwendig „auf DC, OpenAIRE Application Profile, DataCite (und zukünftig dcterms)“ wie Hudak bemerkt (Hudak, Rastislav, Expert*innen-Interview per Mail, Wien, 15.03.2024, s. Anhang 3.2.). Darüber hinaus wurde für den beabsichtigten Umstieg auf RDF auch ein Mapping von UWMETADATA auf MODS gemacht, aber letztendlich nicht vollständig durchgeführt (vgl. Hudak, Rastislav, Expert*innen-Interview per Mail, Wien, 15.03.2024, s. Anhang 3.2.).

Gründe für die erstgenannten Mappings waren die Bedarfe der Partneruniversitäten der Universität Wien und auch die Anforderungen der Projekte, an denen PHAIDRA beteiligt war. Dazu gehörte auch die Zusammenarbeit mit großen Aggregatoren wie OpenAIRE, Europeana und Kulturpool (vgl. Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 2.2.).

Ziel dieser Mappings war es immer, Interoperabilität zu ermöglichen und zu vereinfachen. Jedoch ist für PHAIDRA ein abgeschlossenes Mapping in ein Metadatenformat, das von einem Projektpartner benötigt wird, kein ein für alle Mal abgeschlossener Prozess. Denn auch bei den Projektpartnern kommt es zu Neu- und Weiterentwicklungen und so kann es notwendig sein, Mappings zu überarbeiten oder zu kontrollieren (vgl. Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 2.2.).

In dieser komplexen Konstellation muss immer wieder entschieden werden, mit wem Interoperabilität erreicht werden soll. Dies beinhaltet auch Absprachen zwischen den jeweiligen Organisationen, was nicht immer leicht zu erreichen ist, wie Ganguly bemerkt (vgl. Ganguly, Raman, Expert*innen-Interview, Wien, 20.12.2023, s. Anhang 2.1.). Doch er sieht er auch eine Vereinfachung dieser Aufgabe, denn

> „hier werden wir in den nächsten Jahren sehr starke Änderungen haben, weil wir mehr intelligentere Schnittstellen haben werden, die einfach Transformationen automatisiert machen können. Und auch hier, denke ich, ist der Bereich Machine Learning, also auch wieder in Richtung AI ein ganz wesentlicher Treiber, wo wir einiges verändern können, beziehungsweise erleichtern können“ (Ganguly, Raman, Expert*innen-Interview, Wien, 20.12.2023, s. Anhang 2.1.).

Dies bedeutet, dass in Zukunft notwendige Mappings mit weniger Aufwand auch durch den Einsatz künstlicher Intelligenz ermöglicht werden. Künstliche Intelligenz wird eine Art Adapter bilden können, der auch die Interoperabilität fördern kann.

Ganguly äußert die Befürchtung, dass auch durch den Versuch, Interoperabilität zu stärken, die Diversität der unterschiedlichen eingesetzten Metadatenformate verringert wird, und nur noch einige wenige große Institutionen und Organisationen entscheiden, welche Metadatenformate eingesetzt werden (vgl. Ganguly, Raman, Expert*innen-Interview, Wien, 20.12.2023, s. Anhang 2.1.). Jedoch könnte auch hier der von Ganguly erwähnte Einsatz von Künstlicher Intelligenz von Nutzen sein, da er die unterschiedlichen, vielleicht auch nur von einigen Institutionen oder Organisationen verwendeten Metadatenformate, in die für die Interoperabilität mit anderen Institutionen und Organisationen notwendigen Formate transformieren könnte.

5.2.5. Interoperabilität durch die Verwendung von Harvesting-Protokollen

Harvesting-Protokolle sind eine Möglichkeit Metadaten anderen Institutionen zugänglich zu machen. Für den Einsatz von Harvesting-Protokollen ist es jedoch immer notwendig, auch die richtigen und notwendigen Metadatenformate zur Verfügung zu stellen und, im Fall von PHAIDRA, ein Mapping dorthin zu erstellen.

In PHAIDRA wird die OAI-Schnittstelle und damit auch das Harvesting-Protokoll OAI-PMH von den meisten Repositorien und externen Applikationen verwendet (vgl. Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.). Dort können dann etwa die Metadaten-Schemata oai_dc und oai_openaire abgerufen werden, die aus dem UWMETADATA-Format extrahiert wurden, wie bereits erwähnt. Der OER-Hub verwendet dagegen die Solr-Schnittstelle (vgl. Hudak, Rastislav, Expert*innen-Interview per Mail, Wien, 15.03.2024, s. Anhang 3.2.). Außerdem wird auch eine API-Schnittstelle eingesetzt (vgl. Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 3.1.). Mappings sind daher notwendig, um den Einsatz von Harvesting-Protokollen zu ermöglichen. Generell ist der Einsatz von Harvesting-Protokollen „ein Standard, der in PHAIDRA drin ist, beziehungsweise ein Standard eines jeden Repositories, dass ich meine Metadaten teilen kann“ (Ganguly, Raman, Expert*innen-Interview, Wien, 20.12.2023, s. Anhang 2.1.). Daher ist es auch in unterschiedlichen Szenarien notwendig, ein Mapping von einem Metadatenformat in ein anderes zu erstellen, um Harvesting-Protokolle einsetzen zu können.

5.3. Einsatz von standardisierten und kontrollierten Begriffen

Im Folgenden wird auf Basis der Expert*innen-Interviews betrachtet, welche Rolle der Einsatz von standardisierten und kontrollierten Begriffen für die Beschreibung der digitalen Objekte in digitalen Repositorien, sowie deren Dokumentation, für die Arbeit mit Metadaten spielen.

5.3.1. Die Verwendung von standardisierten und kontrollierten Vokabularen, Thesauren und Taxonomien

Zur Verwendung von standardisierten und kontrollierten Vokabularen, Thesauren und Taxonomien bemerkt Blumesberger:

> „Also wir versuchen immer mehr zu standardisieren. Also das war am Anfang vielleicht nicht ganz so, und das ist jetzt schon ziemlich stark, also wir versuchen wirklich, möglichst standardisiertes Vokabular zu verwenden, zum Teil, was aber nicht wirklich durchhaltbar ist, weil das teilweise nicht passt“ (Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 2.2.).

Bei der Beschreibung der digitalen Objekte in PHAIDRA wird daher großer Wert gelegt auf den Einsatz standardisierter Vokabulare, Thesauren und Taxonomien. Die Standardisierung in diesem Bereich hat seit der Entstehung von PHAIDRA noch an Wichtigkeit gewonnen.

Um den vielfältigen Anforderungen, die von Seiten der Nutzer*innen an PHAIDRA gestellt werden, möglichst gerecht zu werden, werden unterschiedliche kontrollierte Vokabulare in PHAIDRA verwendet, etwa „ÖFOS, Getty, Dewey, BIC, Basisklassifikation,“ und weitere, wie Hudak (Hudak, Rastislav, Expert*innen-Interview per Mail, Wien, 15.03.2024, s. Anhang 3.2.). bemerkt und auch Ganguly erklärt (vgl. Ganguly, Raman, Expert*innen-Interview, Wien, 20.12.2023, s. Anhang 2.1.).

Es kann jedoch Anwendungsfälle geben, für die standardisierte Vokabulare nicht die ideale Lösung bieten, weil die Nutzer*innen des Repositoriums sich durch die Möglichkeiten, die diese standardisierten Vokabulare bieten, nicht repräsentiert fühlen. Als Beispiel nennt Blumesberger das Institut für Kultur- und Sozialanthropologie, das von der Österreichischen Systematik der Wissenschaftszweige (kurz ÖFOS) der Statistik Austria, die ein Pflichtfeld in PHAIDRA ist, nicht repräsentiert fühlt. Es gäbe häufiger

Nutzer*innen, die vor einem ähnlichen Problem stehen (vgl. Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 2.2.). Diese Schwierigkeit tritt auch in anderen Bereichen auf, etwa beim Einsatz der GND, die ebenfalls ein kontrolliertes Vokabular ist. Hier gibt es Nutzer*innen, die andere Bedarfe haben, und sich ein kontrolliertes Vokabular wünschen würden, das noch granularer ist (vgl. Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 2.2.). Es besteht jedoch auch die Möglichkeit, den Vocabulary Server von PHAIDRA zu verwenden und dort bestehende Vokabulare, die den eigenen Anforderungen entsprechen, einzuspielen (vgl. Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 2.2.).

5.3.2. Dokumentation der Verwendung externer als auch lokal entwickelter kontrollierter Vokabulare, Thesauren, Taxonomien

Abgesehen von der Verwendung der kontrollierten Vokabulare, Thesauren und Taxonomien ist auch deren Dokumentation eine Selbstverständlichkeit, wie Blumesberger bemerkt: „Ja, die werden dokumentiert und es gibt auch die Möglichkeit, selbst Vokabularien einzuspielen.“ (Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 2.2.). Auch für Hudak ist es zentral, kontrollierte Vokabulare, Taxonomien und Klassifikationen zu verwenden. Dabei ist es wichtig, auf nachhaltige Einsatzmöglichkeiten zu achten (vgl. Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.).

In die Zukunft blickend beschreibt Hudak auch die Erleichterung in Sachen Dokumentation, die Linked Open Data bringen könnte. Durch den Einsatz von Linked Open Data, das Mapping aufeinander und die Verknüpfung von bestehenden Vokabularen, Taxonomien und Klassifikationen, könnte sichergestellt werden, dass man langfristig mit ihnen arbeiten kann und die Auffindbarkeit ihrer Definitionen und damit auch eine gewisse Dokumentation sichergestellt ist (vgl. Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.). Die stärkere globale Kooperation, die sich automatisch durch die Arbeit mit Linked Open Data einstellen würde, würde daher auch zu einer besseren Dokumentation der verwendeten Vokabulare, Taxonomien und Klassifikationen führen. Denn diese könnten breitere Verwendung erfahren und damit dem Veralten und Verschwinden entgegenwirken.

5.4. Festlegung der rechtlichen Nutzungsbedingungen

In den nächsten Abschnitten wird auf Basis der Expert*innen-Interviews betrachtet, wie die Festlegung der rechtlichen Nutzungsbedingungen in PHAIDRA gestaltet ist und welche Auswirkungen dies auf die Arbeit mit Metadaten hat. Dabei wird die Beschreibung der Rechte, ihr Management und ihre Kommunikation sowie die Verwaltung rechtlicher Zugangsbedingungen beleuchtet.

5.4.1. Beschreibung der Rechte

Die Beschreibung der Rechte in Bezug auf die digitalen Objekte in PHAIDRA beginnt schon vor dem Upload der digitalen Objekte. Hier gibt es „Nutzungsbedingungen, die akzeptiert werden müssen, bevor man etwas hochladen kann. Und dort steht drinnen, dass man eben nichts hochlädt, was irgendwelchen Rechten widerspricht. Das heißt, man sagt, man hat die Rechte auf dieses Objekt. Damit ist einmal die Person, die hochlädt, verantwortlich für das, was diese Person hochlädt.“ (Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 2.2.). Das bedeutet, die Person, die etwas hochlädt, muss im Vorfeld alle rechtlichen Fragen in Zusammenhang mit dem jeweiligen Objekt klären.

In einem zweiten Schritt, bemerkt Blumesberger (vgl. Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 2.2.), kann dann über die Weiterverwendung der Objekte entschieden werden. Dafür gibt es zwei grundlegende Möglichkeiten: Es besteht die Option, Objekte ganz zu sperren, damit sind sie für die Öffentlichkeit nicht sichtbar. Dann gibt es noch die Alternative, Objekte zugänglich zu machen und ihre Verwendung vom Urheberrecht beschränken zu lassen. So wäre in diesem Fall etwa ein Download möglich, aber keine Weiterverarbeitung. Außerdem kann die zukünftige erlaubte Arbeit mit öffentlich zugänglichen digitalen Objekten in PHAIDRA auch durch die Creative Commons Lizenzen gesteuert werden. Auch Hudak beschreibt diese zwei Schritte (vgl. Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.). Einmal vergebene Lizenzen können jedoch nicht mehr geändert werden (vgl. Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 2.2.).

5.4.2. Management und Kommunikation von Rechten

Für das Management und die Kommunikation von Rechten sieht PHAIDRA vor, dass diese auch maschinell austauschbar sind: „die Angaben über die Rechte werden an Harvester weitergegeben, beispielsweise über die API-Schnittstelle“ (Blumesberger, Susanne, Expert*innen-Interview, Wien, 13.03.2024, s. Anhang 3.1.). Darüber hinaus werden die Lizenzen mittels eines Uniform Resource Identifiers (URI) kommuniziert und sind damit von Maschinen lesbar. Wird keine Lizenz beim Upload in PHAIDRA vergeben, so wird dies auch mit einer URI kommuniziert, die auf das trotzdem geltende Urheberrecht verweist (vgl. Hudak, Rastislav, Expert*innen-Interview per Mail, Wien, 15.03.2024, s. Anhang 3.2.).

5.4.3. Verwaltung rechtlich begründeter Zugangsbedingungen zu Objekten

Um rechtlich begründete Zugangsbedingungen zu den Objekten in PHAIDRA zu verwalten, bemerkt Blumesberger: „Einschränkungen der Sichtbarkeit in PHAIDRA gibt es, dahinter liegt die Struktur der Universität, man kann auf die ganze Uni, auf Institute, auf Gruppen von Personen, die selbst definiert werden können und auch auf Einzelpersonen einschränken.“ (Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 3.1.). Damit ist sichergestellt, dass Objekte nur archiviert werden können, ohne öffentlich zugänglich zu sein. Außerdem können weitere Zugangsbeschränkungen getroffen werden, um etwa die Objekte nur bestimmten Personen zugänglich zu machen, wie auch Hudak bemerkt (vgl. Hudak, Rastislav, Expert*innen-Interview per Mail, Wien, 15.03.2024, s. Anhang 3.2.). Jedoch bleiben die Metadaten trotz der Sperre der Objekte sichtbar (vgl. Hudak, Rastislav, Expert*innen-Interview per Mail, Wien, 15.03.2024, s. Anhang 3.2.). Wenn Objekte hochgeladen werden sollten, für die die Rechte nicht vorhanden sind, können diese auch wieder gelöscht werden.

5.5. Metadaten für langfristigen Erhalt und Pflege der digitalen Objekte

In den kommenden Abschnitten wird betrachtet, wie der langfristige Erhalt und die Pflege von digitalen Objekten durch Metadaten in PHAIDRA sichergestellt werden. Wie wichtig dies ist, zeigt dieses Zitat von Hudak: „Die große Sorge ist, dass ich am Ende irgendein Data Lake habe oder Datenfriedhof am besten. Und um das zu vermeiden, brauche ich zu wissen, welche Dateien was sind, und dafür sind die Metadaten da“ (Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.). Die Metadaten werden dabei in Kategorien unterteilt, die nun im Hinblick auf ihre Verwendung in PHAIDRA betrachtet werden.

5.5.1. Der Einsatz von administrativen Metadaten

Zunächst gibt es administrative Metadaten, die werden „oft nicht in diesen bibliografischen Metadaten angezeigt, aber sind drin im System, da wo dann steht, wann das hochgeladen wurde, wann das editiert wurde. Und (…) das ist auch natürlich wichtig“ (Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.). Zu den administrativen Metadaten kann etwa der Account (in PHAIDRA die Metadatenkategorie ownerid) gehören, von dem aus die digitalen Objekte hochgeladen werden. Diese Information legt fest, wer die Erlaubnis hat, die Objekte zu bearbeiten. Darüber hinaus ist aus dieser Angabe auch ersichtlich, wer den Nutzungsbedingungen zugestimmt hat. Auch die Metadatenkategorie „audit log“ gehört zu den administrativen Metadaten. Sie ist in der Lage festzuhalten, wer wann und wie das digitale Objekt oder die Metadaten bearbeitet hat. Darüber hinaus, bemerkt Hudak, auch bibliographische Metadaten seien administrative Metadaten, die dazu verwendet werden, wie etwa mittels der Lizenzangaben die weitere Nutzung der digitalen Objekte festgelegt werden kann. Dies ist nicht nur für potentielle Nutzer*innen relevant, sondern auch für Überlegungen innerhalb der Bibliothek oder des Repository-Teams, wenn es darum geht, ob die digitalen Objekte über Harvester weiterverbreitet werden sollen (vgl. Hudak, Rastislav, Expert*innen-Interview per Mail, Wien, 15.03.2024, s. Anhang 3.2.). Administrative Metadaten spielen daher eine wichtige Rolle

bei der Beschreibung der digitalen Objekte in PHAIDRA und anderen digitalen Repositorien.

5.5.2. Der Einsatz von technischen Metadaten

Auch technische Metadaten werden in PHAIDRA eingesetzt. Dazu gehört etwa die Metadatenkategorie Größe zur jeweiligen Datei, aber auch die Metadatenkategorie MIME type (vgl. Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.). Diese beschreibt, ob es sich um eine Bilddatei oder ein PDF handelt und ermöglicht Repositorien generell und damit auch PHAIDRA, zu entscheiden, wie etwa das Objekt angezeigt werden kann (vgl. Hudak, Rastislav, Expert*innen-Interview per Mail, Wien, 15.03.2024, s. Anhang 3.2.). Allerdings bemerkt Hudak dazu: „[I]ch finde manchmal übertreibt man ein bisschen mit den technischen Metadaten. Ich brauche in diesen Metadaten nicht die Information darüber, wie groß die Datei ist, weil die Datei ist irgendwie groß und das kann ich immer herausfinden, ja? Was in der Datei drinnen ist, quasi welches MIME type das hat, das kann ich nicht so leicht herausfinden, das ist dann zum Beispiel schon hilfreich.“ (Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.). Daher haben nicht alle Arten von technischen Metadaten die gleiche Relevanz.

5.5.3. Der Einsatz von Erhaltungsmetadaten

Auch Erhaltungsmetadaten werden in PHAIDRA eingesetzt, um die langfristige Pflege und den Erhalt der digitalen Objekte sicherzustellen. Dazu gehört etwa die Metadatenkategorie checksum, die dazu dient, die Verschlechterung der Daten von digitalen Objekten zu verhindern oder ungewollte Änderungen zu unterbinden. Es werden darüber hinaus auch andere Erhaltungsmetadaten in PHAIDRA eingesetzt, so auch, um den Entstehungsprozess eines digitalen Objektes zu dokumentieren, etwa bei der Digitalisierung einer Fotografie (vgl. Hudak, Rastislav, Expert*innen-Interview per Mail, Wien, 15.03.2024, s. Anhang 3.2.).

5.5.4. Der Einsatz von Recordkeeping-Metadaten für das Objekte-Management

Laut Hudak sind Recordkeeping-Metadaten eher im Bereich der Arbeit von Archiven zu finden. Dieser Bereich hat jedoch eine Schnittmenge mit den Metadatenkategorien, die in digitalen Repositorien zu den administrativen, technischen, strukturellen oder auch Erhaltungs-Metadaten gezählt werden (vgl. Hudak, Rastislav, Expert*innen-Interview per Mail, Wien, 15.03.2024, s. Anhang 3.2.).

5.5.5. Der Einsatz von strukturellen Metadaten

Strukturelle Metadaten werden in PHAIDRA ebenfalls eingesetzt. Sie dienen dazu, die digitalen Objekte miteinander in Beziehung zu setzen. Etwa wenn es darum geht, hierarchische Strukturen abzubilden, Versionierungen von Objekten oder Verbindungen zwischen Objekten darzustellen (vgl. Hudak, Rastislav, Expert*innen-Interview per Mail, Wien, 15.03.2024, s. Anhang 3.2.). Dazu kann es auch gehören, Forschungsdaten mit Publikationen in Beziehung zu setzen. Aufgabe der strukturellen Metadaten ist es, diese Beziehungen abzubilden und nachvollziehbar zu machen (vgl. Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.).

5.6. Metadaten als digitale Objekte – Einsatz eines persistenten, eindeutigen Identifikators

Versteht man Metadaten als digitale Objekte, so sollten sie auch den Anforderungen, die an digitale Objekte gestellt werden, gerecht werden. Einer der Aspekte, der dabei bisher nicht betrachtet wurde, ist die Verwendung eines persistenten, eindeutigen Identifikators für Metadaten, seine Notwendigkeit und der Einfluss, den er auf die Arbeit mit Metadaten und eventuelle Mappings haben könnte.

Bei der Frage, ob Metadaten als Gesamtes auch persistente, eindeutige Identifikatoren haben sollten, ist dies für Blumesberger nicht sinnvoll, da die Metadaten in PHAIDRA auch verändert werden können, wobei diese Veränderung dokumentiert wird. Darüber hinaus verweisen die persistenten Identifikatoren der digitalen Objekte auf eine

Landing-Page, auf der auch die Metadaten zu finden sind. Daher ist es nicht notwendig, einen eigenen Identifikator für die gesamten Metadaten zu haben (vgl. Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 2.2.). Jedoch bietet die Arbeit mit RDF die Möglichkeit, für die einzelnen Metadatenelemente Identifikatoren zu verwenden: Dies sind „Links und Identifikatoren, und die sind global und unique und persistent. Das heißt, zeigen nicht morgen auf etwas Anderes und dadurch mache ich die Metadaten sozusagen automatisch verständlich für alle anderen, auf diese Weise. Das kann man mit normalen Standards aber nicht wirklich erreichen“ (Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.).

5.7. Die Modellierung von Metadaten: Metadatenkategorien und die Beziehungen zwischen ihnen

Bei der Modellierung von Metadaten für einen bestimmten Bereich, spielen neben den verwendeten Metadatenkategorien auch die zwischen ihnen bestehenden Beziehungen eine wichtige Rolle, die sich auch auf die Arbeit mit Metadaten auswirken können.

Das digitale Repositorium PHAIDRA der Universität Wien basiert auf der Software Fedora Commons. Fedora Commons wurde immer wieder weiterentwickelt (vgl. Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.). Dabei wurde an einem bestimmten Punkt „von Version 3 auf 4 auf eine andere Metadaten-Struktur, eigentlich ein komplett anderes Konzept, gegangen“ (Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.). Es kam also zu einer tiefgreifenden Änderung in der Art, wie die Metadaten strukturiert waren und damit veränderte sich das Modell, auf dem die Metadaten basierten.

Hudak beschreibt diesen Wechsel wie folgt: „Die Fedora 3, auf der wir noch immer sind, war quasi XML-basiert und da hat das UW-Metadatenformat sehr gut reingepasst. In der Version 4 war es eine sogenannte Linked-Data-Plattform plötzlich. Und man hat RDF erwartet und / Man kann nach wie vor mit Metadaten als einem einfachen XML-File damit arbeiten. Machen auch sicher viele, weil dieser Umstieg auf RDF ist nicht trivial.“ (Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.).

Diese Veränderung in der unterstützten oder auch notwendigen Metadaten-Struktur für die Arbeit mit Fedora Commons bestand in einem Wechsel von einer XML-Struktur, mit der das Format UWMETADATA gut funktionierte, hin zu einer RDF-Struktur, die auf der Idee von Linked Data basiert. Das machte einen Übergang von der XML-Struktur zu einer RDF-Struktur und damit auch ein Mapping notwendig. Blumesberger erwähnt ebenfalls die Entwicklung der neuen Fedora Version und die Notwendigkeit des Umstiegs auf die neue Metadaten-Struktur (vgl. Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 2.2.). Auch Ganguly beschreibt den Wechsel auf Linked Data mit einer neuen Fedora-Version, der eine Migration auf die neue Metadaten-Struktur notwendig macht (vgl. Ganguly, Raman, Expert*innen-Interview, Wien, 20.12.2023, s. Anhang 2.1.).

Die strukturellen Grundlagen von RDF selbst beschreibt Hudak wie folgt: „Also das RDF basiert auf diesem total primitiven Kern von Grammatik, Subject Predicate Object, und mit dem baue ich den Graph“ (Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.). Ein zu diesem Zeitpunkt bestehender Weg, um auf die gewünschte RDF-Struktur zu kommen, war zunächst auf das Format MODS zu mappen, wie es die Samvera-Community getan hatte. Diese arbeitete mit MODS-Metadaten in Fedora und hatte ein Mapping von MODS nach RDF gemacht, um auf die Version 4 von Fedora umsteigen zu können (vgl. Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.). Das PHAIDRA-Team wollte diesem Weg folgen und da bereits erste Erfahrungen mit dem MODS-Format vorhanden waren (vgl. Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.), entschied man sich zu einem Mapping von UW-Metadaten (einem XML-Format) auf MODS (auch ein XML-Format), um darüber und durch die Erfahrungen der Samvera-Community zum RDF-Format zu kommen (vgl. Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.). Dieses Mapping sollte daher den Umstieg auf die neue Datenstruktur und damit ein neues Datenmodell ermöglichen, war dabei allerdings nur ein Zwischenschritt, der rückblickend vor allem der Orientierung und Inspiration diente, im Hinblick auf die Metadatenkategorien, die man im neuen Metadatenformat verwenden wollte (vgl. Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.). Dies ergänzt auch die Bemerkung von Bellotto und Bettella, die das Vermitteln zwischen unterschiedlichen

Schemata, ihre Strukturen, Modellierungen und Repräsentationen als Prozess sehen, der auch dazu führen kann, die eigenen Metadatenstrukturen weiter zu verbessern (vgl. Bellotto & Bettella 2019:177ff.). Bei PHAIDRA an der Universität Wien reicht dieser Prozess in zukünftige Entwicklungen hinein.

Die Vorteile, die sich durch das Mapping auf MODS und dann auf eine Metadaten-Struktur und -Darstellung ergeben, die von der neuen Fedora-Version unterstützt werden, findet sich in der Darstellung der Metadaten und ihrer möglichen Verwendung: RDF ist maschinenlesbar und die Maschinenlesbarkeit und der Einsatz von RDF erleichtert wieder den Austausch von Metadaten (vgl. Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.). Der Linked Data-Ansatz und damit auch RDF haben daher Merkmale, die neue Möglichkeiten erschließen, wie etwa die globale Zugänglichkeit, Eindeutigkeit und Stabilität der verwendeten Elemente. Dazu gehört auch die Möglichkeit, die Beziehungen zwischen den einzelnen Metadaten-Werten zu analysieren und so logische Schlüsse zu ziehen (Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.).

In diesem Kapitel wurden die durchgeführten Expert*innen-Interviews auf Basis der Struktur des erstellten Kategorienleitfadens ausgewertet. Im folgenden Kapitel werden die Ergebnisse diskutiert und, wo möglich, in Verbindung zum Forschungsstand gebracht.

6. Diskussion

Diese Kapitel betrachtet und interpretiert die Ergebnisse der Expert*innen-Interviews im Hinblick auf die Forschungsfragen und stellt den Zusammenhang zum Forschungsstand her.

6.1. Diskussion zur Verwendung von Community Standards

Bei den Betrachtungen zur Verwendung von Community Standards ist deutlich geworden, dass bei der Verwendung von bestimmten Metadaten-Schemata in digitalen Repositorien zwei Aspekte wichtig sind: Einerseits muss man die Bedarfe der unterschiedlichen Nutzer*innengruppen, etwa Vertreter*innen verschiedener Fachgebiete, eines digitalen Repositoriums durch ein geeignetes Metadatenschema berücksichtigen. An der Universität Wien wurde dafür das Schema UWMETADATA entwickelt. Gleichzeitig ist es erforderlich Community Standards zu verwenden, um an unterschiedlichen Projekten teilnehmen zu können, die Metadaten anderen Organisationen zugänglich zu machen und Interoperabilität zu ermöglichen. Um zu diesen Community Standards zu gelangen, sind Mappings des eigenen Formats, im Fall des digitalen Repositoriums PHAIDRA von UWMETADATA, auf die Community Standards notwendig gewesen.

Damit wird deutlich, dass die Rolle von Standards für die Zugänglichkeit, Wiederverwendung und Interoperabilität von Metadaten in digitalen Repositorien zwar zentral ist, wie auch Bellotto und Bettella festhalten (vgl. 2019:177ff), darüber hinaus jedoch nicht vergessen werden darf, dass ein digitales Repositorium für eine Gemeinschaft und aus ihr entsteht und ihren Bedarfen auch gerecht werden muss.

Der zweite Aspekt, die Verwendung von Application Profiles, ergänzt auch die Perspektive von Chan und Zeng (2006a; 2006b), die untersuchten, welche Mittel eingesetzt werden können, um eine Suche über mehrere Repositorien durchzuführen. Zentral ist für sie dabei die Herstellung von Interoperabilität, die durch individuelle Schemata oder die Verwendung von Application Profiles (wie auch in PHAIDRA) nicht vorhanden ist, sowie durch Zusammenarbeit zwischen unterschiedlichen Datenbanken notwendig

wird. Diese Ursachen für Mappings zeigen sich auch bei der Arbeit mit einzelnen digitalen Repositorien, die etwa an Kooperationen teilnehmen oder ihre Metadaten anderen Organisationen zur Verfügung stellen wollen und dafür Interoperabilität herstellen müssen.

Eine Kombination der Vorteile der Verwendung von UWMETADATA (die Beschreibung entsprechend den Bedarfen der Nutzer*innen) und der Verwendung von Community Standards bietet in Zukunft der Einsatz von RDF, das eine genaue Beschreibung durch die Verwendung von global verbreiteten Standards ermöglicht. Doch auch hier sollte man sich nicht auf ein eng abgegrenztes Feld beschränken, da es sonst zu ähnlichen Problemen kommen kann wie bei eigens entwickelten Metadatenformaten und wieder Mappings notwendig werden können. Daher sollte man, wie auch Hudak bemerkte, die Möglichkeiten ausschöpfen, die RDF bietet: Es kann sehr leicht erweitert werden und ist dadurch in der Lage, auf unterschiedliche Bedarfe flexibel zu reagieren. Dadurch kann in komplexen Situationen Interoperabilität herstellen werden, wie auch Bellotto und Bettella (vgl. Bellotto & Bettella 2019:182f.) sowie Fiala und Huggle (vgl. Fiala & Huggle 2019:2) anmerken. Darüber hinaus kann im Allgemeinen auch bei der Verwendung desselben Standards dieser von unterschiedlichen Institutionen – etwa beim Einsatz von Application Profiles – unterschiedlich interpretiert werden, was dann beim Versuch Interoperabilität herzustellen wieder Mappings notwendig machen kann, wie oben auch bei Chan und Zeng (2006a; 2006b) für Suchen über mehrere Repositorien hinweg festgehalten.

Doch die Möglichkeit digitale Objekte detailliert zu beschreiben, ist nicht der einzige Vorteil der Arbeit mit global verwendeten Standards in RDF. Durch die Verwendung dieser unterschiedlichen Standards oder Ontologien in RDF werden die Metadaten von PHAIDRA in einen globalen Kontext aus Standards eingebettet und ermöglichen dadurch Interoperabilität in einem um einiges weitreichenderen Umfang als es bisher möglich war.

6.2. Diskussion zur Unterstützung von Interoperabilität

Im Hinblick auf die Vergabe von Metadaten ist deutlich geworden, wie wichtig eine detaillierte Beschreibung der digitalen Objekte durch detaillierte Metadaten ist. Sie ist eine unbedingte Voraussetzung, um den unterschiedlichen Ansprüchen an die Metadatenformate eines Repositoriums einer vielfältigen Universität mit zahlreichen Disziplinen – wie etwa der Universität Wien – gerecht zu werden. Jedoch werden an diese Beschreibung häufig die unterschiedlichsten, oft fachbereichsspezifischen Ansprüche gestellt. Um diesen Anforderungen gerecht zu werden, reicht es nicht aus, die Metadaten auf der beschränkten Anzahl von Metadatenfeldern eines weiter verbreiteten Standards basieren zu lassen, um etwa Interoperabilität zu ermöglichen. Bei der Entwicklung von PHAIDRA wurde daher von Beginn an UWMETADATA verwendet, das in der Lage ist, auch komplexere Beschreibungen durch detaillierte Metadaten abzubilden. Dazu gehört etwa die Darstellung mehrerer Ebenen eines Digitalisats. Die sehr detaillierte Beschreibung, die manchmal von Forschenden gewünscht wird, stellt das PHAIDRA-Team immer wieder vor neue Aufgaben. Durch das Format UWMETADATA können die digitalen Objekte sehr genau beschrieben werden und bleiben auch in anderen Kontexten verständlich. Doch würde dieses Format bei der Weitergabe der Metadaten diese Qualität nur begrenzt behalten, da es rein technisch nicht immer interoperabel wäre. Für die technische Interoperabilität werden Mappings in andere Metadatenformate eingesetzt. Um jedoch gleichzeitig eine detaillierte Beschreibung, den Einsatz von global verwendeten Standards, Interoperabilität und Maschinenlesbarkeit zu ermöglichen, soll in Zukunft RDF eingesetzt werden. Damit können also detaillierte Metadaten vergeben werden, die ausreichend Informationen zu den digitalen Objekten bieten, damit diese auch außerhalb des Kontextes der Universität Wien und PHAIDRA richtig gedeutet werden können. Darüber hinaus ist es mit dem Einsatz von RDF auch möglich, die Metadaten in ein globales Netz von Daten einzubetten. Der Weg zu RDF könnte rein theoretisch über ein Mapping führen, welches jedoch sehr komplex wäre. Darüber hinaus soll auch die detaillierte Beschreibung der Metadaten in Zukunft aus Perspektive mehrerer Fachbereiche ermöglicht werden, um Interdisziplinarität zu fördern. Dabei könnte es auch in Zukunft wichtig sein,

die verschiedenen Blickpunkte der Beschreibung wieder miteinander in Beziehung zu setzen und damit auch aufeinander zu mappen.

Für die Zusammenarbeit mit Aggregatoren, die normalerweise Harvesting-Protokolle einsetzen, wurden die Metadaten aus dem Format UWMETADATA in ein von den Aggregatoren/Harvestern benötigtes Format, wie etwa oai_dc und oai_openaire, gemappt. Die Mappings in die unterschiedlichen Metadatenformate sind für die Kooperation mit Aggregatoren notwendig und ermöglichen diese. Darüber hinaus sind sie bei der Zusammenarbeit in unterschiedlichen Projekten wichtig. Außerdem kann es selbst bei der Kollaboration mit langjährigen Partnern notwendig werden, ein bestehendes Mapping zu analysieren und zu überarbeiten, da es immer wieder zu Weiterentwicklungen und Neuerungen kommen kann. In Zukunft werden dafür wohl auch maschinell erstellte Mappings eingesetzt werden können, was dazu beitragen könnte, die Vielfalt an Metadatenformaten zu bewahren und dabei gleichzeitig Interoperabilität zu fördern.

Was die Zugänglichmachung der Metadaten-Schemata betrifft, wurde von UWMETADATA keine öffentliche Dokumentation für Interessierte erstellt. Jedoch sind die verwendeten Community Standards gut dokumentiert, meist auf eigenen Webseiten. Eine gute und leicht zugängliche Dokumentation des Metadatenformats ist auch ein Vorteil, der sich aus der Verwendung von RDF ergeben würde.

Außerdem lässt sich sagen, dass in der Arbeit mit PHAIDRA Interoperabilität durch den Einsatz von Mappings in weitere Metadatenformaten mit anderen Institutionen und Organisationen erzielt wurde und in Zukunft durch die Einbindung von Künstlicher Intelligenz noch erleichtert wird. Gleichzeitig können einzelne Mappings nicht als Abschlusspunkt gesehen werden, da es immer wieder neue Szenarien geben kann, die weitere Mappings notwendig machen.

6.3. Diskussion zum Einsatz von standardisierten und kontrollierten Begriffen

Bei der Arbeit mit PHAIDRA wird generell sehr darauf geachtet standardisierte und kontrollierten Vokabulare einzusetzen. Darüber hinaus ist es für die Nutzer*innen möglich

standardisierte Vokabulare, die es in PHAIDRA noch nicht gibt, über den PHAIDRA Vocabulary Server zu verwenden. Dabei scheint der Einsatz von standardisierten und kontrollierten Vokabularen, Thesauren und Taxonomien zunächst keinen Einfluss auf die Notwendigkeit von Mappings zu haben. Ruft man sich jedoch die Aussage von Ganguly (vgl. Ganguly, Raman, Expert*innen-Interview, Wien, 20.12.2023, s. Anhang 2.1.) in Kapitel 5.2.1. in Erinnerung, der die Notwendigkeit von Mehrfachbeschreibungen der digitalen Objekte durch die Nutzer*innen unterschiedlicher Fachbereiche als eine Möglichkeit sah, Interdisziplinarität herzustellen, könnte es doch notwendig sein, standardisierte und kontrollierte Vokabulare, Thesauri und Taxonomien aufeinander zu mappen.

Zusammenfassend lässt sich sagen, die Verwendung von kontrollierten und standardisierten Vokabularen, Thesauren und Taxonomien sowie deren Dokumentation ist ein wichtiger Aspekt in der Arbeit mit PHAIDRA. Bei der Arbeit mit PHAIDRA selbst scheint dieser Aspekt bisher noch keine Mappings von einem Vokabular, einem Thesaurus, einer Taxonomie notwendig gemacht zu haben. Jedoch könnte es sein, dass es irgendwann einmal notwendig sein wird, dies zu tun: Einerseits, um bei Mehrfachbeschreibungen durch Nutzer*innen unterschiedlicher Fachbereiche Verbindungen zwischen den Beschreibungen und damit Interdisziplinarität zu ermöglichen. Darüber hinaus könnte ein Mapping notwendig werden, etwa wenn ein bestimmtes Vokabular, ein bestimmter Thesaurus oder eine bestimmte Taxonomie nicht mehr gepflegt wird und veraltet.

6.4. Diskussion zur Festlegung der rechtlichen Nutzungsbedingungen

Es wurde auch die Herangehensweise an die rechtlichen Nutzungsbedingungen in PHAIDRA betrachtet. Diese umfassen die Zustimmung zu den Nutzungsbedingungen, die Vergabe einer Lizenz für die erlaubte Art der Weiterverwendung der digitalen Objekte, die maschinelle Austauschbarkeit der Lizenzen, sowie die auf unterschiedlichen Ebenen mögliche Beschränkung des Zugangs zu den digitalen Objekten. Dieser Aspekt der Arbeit mit Metadaten hat bei der Arbeit mit PHAIDRA keinen Einfluss auf die Notwendigkeit von Mappings gezeigt.

6.5. Diskussion zur Verwendung von Metadaten für langfristigen Erhalt und Pflege der digitalen Objekte

Die Untersuchung von unterschiedlichen Arten von Metadaten für die langfristige Pflege von digitalen Objekten machte deutlich, dass sie im Hinblick auf die Notwendigkeit von Mappings und deren Vorteile keine Rolle spielen. Nichtsdestotrotz spielen sie eine wichtige Rolle bei der Beschreibung der digitalen Objekte und erleichtern, wenn sie gut gewählt und vergeben werden, die zukünftige Verwendung von digitalen Objekten und auch die Erstellung von Mappings.

6.6. Diskussion zur Verwendung von Metadaten als digitale Objekte – Einsatz eines persistenten, eindeutigen Identifikators

Bei der Frage, ob es sinnvoll wäre, einen persistenten eindeutigen Identifikator für die gesamten, bereits erarbeiteten Metadaten eines digitalen Objekts zu vergeben, wurde deutlich, dass dies nicht zielführen ist. Schließlich können die einmal erstellten Metadaten in PHAIDRA verändert werden. Betrachtet man die Frage jedoch aus einer leicht abgeänderten Perspektive und blickt auf die einzelnen Metadatenelemente, ist es sinnvoll, diese mittels eines Identifikators eindeutig zu definieren, wie es etwa im Rahmen der Nutzung von RDF durch URIs möglich ist. Hier wäre es jedoch gut, weitere Untersuchungen dazu anzustellen.

6.7. Diskussion zur Modellierung von Metadaten: Metadatenkategorien und die Beziehungen zwischen ihnen

Anhand des digitalen Repositoriums PHAIDRA der Universität Wien wurde auch betrachtet, wie die Veränderung des dem Repositorium zugrunde liegenden Metadaten-Modells sich auf die Notwendigkeit von Mappings auswirken kann. Dabei zeigte sich, dass der Wechsel von einer Metadaten-Struktur der dem digitalen Repositorium zugrunde liegenden Software auf eine andere ein Mapping notwendig machen kann, wie auch Bellotto

und Bettella für den Einsatz des Repositoriums PHAIDRA an der Universität Padua festhalten (vgl. Bellotto & Bettella 2019:182f.). Im Fall von PHAIDRA an der Universität Wien (sowie Padua) und dem Wechsel zu einem Linked Data-Datenmodell war das Mapping auf MODS jedoch nur ein Zwischenschritt auf dem Weg zum neuen Metadaten-Modell, der rückblickend der Inspiration und Orientierung diente und durch weitere Arbeit ergänzt wurde. Der Vorteil, der sich durch die Verwendung des neuen RDF-Modells ergeben hätte und über einen anderen Weg ergibt, sind Maschinenlesbarkeit der vollständigen Metadaten, die bisher im Format UW-Metadaten vorliegen. Darüber hinaus werden durch das neue Metadatenmodell die vorhandenen (vielfältigen) UW-Metadatenelemente in eindeutige, global verwendete und voraussichtlich langfristig stabile Standards eingebettet, die eine nachhaltige Nutzung sicherstellen. Außerdem wären durch eine Verwendung von RDF Maschinen in der Lage, Schlüsse auf Basis der Metadatenstruktur zu den vorhandenen Daten zu ziehen.

Diese Entwicklung stimmt auch mit Marcondes (2013:7) überein, der fordert, dass Wissensorganisationssysteme, zu denen auch Metadaten gehören, sich die Vorteile, die Ontologien bieten zunutze machen, eben weil letztere maschinelles Schließen ermöglichen. Diese Vorteile bestehen für die Institution an sich und auch in der Kooperation mit anderen Institutionen: etwa in dem durch das neue Format die Metadaten mehr Nutzer*innen zugänglich gemacht werden, wie Southwick für die Umwandlung der Metadaten einer digitalen Sammlung der Universität Nevada in Las Vegas nach Linked Open Data festhält (vgl. Southwick 2015:2).

7. Schlussbemerkungen

Diese Arbeit hat sich mit Metadaten in digitalen Repositorien beschäftigt. Zentral waren dabei die Forschungsfragen:

- Was kann Mappings zwischen unterschiedlichen Metadatenformaten in digitalen Repositorien notwendig machen?
- Welche Vorteile können sich durch Mappings in andere Metadatenformate ergeben?

Um sie beantworten zu können, wurden teilstrukturierte Leitfaden-Interviews mit drei Expert*innen geführt, die an der Arbeit rund um PHAIDRA, dem digitalen Repositorium der Universität Wien, beteiligt sind. Die Interviews wurden mittels der Methode der qualitativen Inhaltsanalyse, genauer der deduktiven Kategorienanwendung nach Mayring (2022) ausgewertet. Ziel war es, durch die Interviews Informationen zu generieren und Erkenntnisse zu gewinnen, die eventuell auf weitere digitale Repositorien und Mapping-Prozesse generell übertragbar sein könnten.

Aus den Interviews ergaben sich mehrere Gründe, die Mappings von einem Metadatenformat in ein anderes notwendig machten. Die Universität Wien hat mit dem Metadatenformat UWMETADATA ein komplexes und detailliertes Format, das den unterschiedlichsten Ansprüchen seiner zahlreichen verschiedenen Nutzer*innen-Gruppen gerecht werden kann. Zunächst sind daher Mappings in andere Metadatenformate und Metadaten-Standards notwendig, um die eigenen Metadaten anderen Organisationen und Institutionen zugänglich zu machen und in Projekten kooperieren zu können. Etwa wenn die Metadaten in einem komplexen und detaillierten Format vorliegen, das vor allem den unterschiedlichsten Ansprüchen der eigenen zahlreichen und verschiedenen Nutzer*innen-Gruppen gerecht werden soll. Hier ist es oft notwendig, technische Interoperabilität durch Mappings in ein anderes Metadatenformat herzustellen. Außerdem ist es auch bei der Verwendung von verbreiteten Standards möglich, diese für die eigenen Zwecke zu interpretieren, etwa durch Application Profiles. Dies kann bei der Zusammenarbeit mit anderen Institutionen wieder Mappings in ein gemeinsames Verständnis eines Metadatenformates notwendig machen. Außerdem kann es immer wieder bei einzelnen Institutionen oder Projektpartnern zu Weiterentwicklungen und Neuerungen im Bereich der Metadaten kommen, sodass dies auch immer wieder Mappings notwendig machen kann.

Auch die Zusammenarbeit mit Aggregatoren und der Einsatz von Harvesting-Protokollen kann Mappings notwendig machen, damit die eigenen Metadaten von anderen weiterverwendet werden können.

Ein Grund, der ebenfalls Mappings von einem Metadatenformat in ein anderes notwendig machen kann, ist die Änderung des der Repositorium-Software zugrunde liegenden Datenmodells. Im spezifischen Fall von einer XML-Struktur zu einer RDF-Struktur. Selbst wenn dieses Mapping dann nur eine Näherung ist auf dem Weg zum neuen Datenmodell und vor allem Erkenntnisgewinn und Inspiration für die weitere Arbeit bringt, ist es bedingt durch diese technische Veränderung, die auf globaler Ebene und in vielen weiteren Anwendungsbereichen wirkt. Metadaten-Mappings sind damit immer wieder notwendig, um gute Metadaten sicherzustellen, aber auch um technischen Veränderungen Rechnung zu tragen.

Die Vorteile, die sich durch Mappings von einem Metadatenformat in ein anderes ergeben, beginnen allgemein bei besseren Kooperationsmöglichkeiten durch die Verwendung gemeinsamer Standards oder Metadatenformate. Im spezifischen Fall der Verwendung von RDF würde dadurch der Einsatz von global verwendeten, durch persistente Identifikatoren sehr stabilen und darüber hinaus maschinenlesbaren Standards möglich. Gleichzeitig ermöglicht dies eine detaillierte Beschreibung der Objekte, wird damit den Bedarfen der unterschiedlichen Institutionen und Organisationen gerecht und fördert eine nachhaltige Arbeit mit den Metadaten.

Auch wenn diese Arbeit ein paar Aspekte des Forschungsstandes ergänzen konnte, muss festgehalten werden, dass es sich lediglich um eine Studie anhand der Metadatenformate des digitalen Repositoriums PHAIDRA an der Universität Wien, mit den für das Repositorium und die Universität spezifischen Umständen handelt. Untersuchungen zu anderen digitalen Repositorien, die andere Nutzer*innen und andere Ziele haben oder anderen Arten von Organisationen dienen, könnten möglicherweise zu anderen oder zusätzlichen Ergebnissen führen. Ein Wunsch für weiterführende Forschung wäre daher, weitere Studien zu unterschiedlichen Arten von digitalen Repositorien durchzuführen, um die Ergebnisse zu überprüfen und gegebenenfalls zu ergänzen.

Anhang

Anhang 1: Interviewleitfaden für die Expert*innen-Interviews

Einführung
• Zunächst möchte ich Sie bitten, sich kurz vorzustellen: Wie heißen Sie, wo arbeiten Sie und was genau ist ihr Aufgabenbereich?
• Wie wichtig ist in Ihrer Arbeit die Arbeit mit digitalen Repositorien und Metadaten?
• In welcher Form haben Sie an der Arbeit an PHAIDRA und auch den dort verwendeten Metadatenformaten mitgewirkt?

Themenblock 1: PHAIDRA und seine Metadaten-Schemata
• PHAIDRA ist inzwischen schon etwa 15 Jahre alt. Wie hat sich die Arbeit mit den digitalen Objekten und Metadaten, ihrer Vergabe und Darstellung in dieser Zeit entwickelt? o Welche Phasen gab es da? Was wurde in diesen Phasen erreicht? o Gab es in den letzten Jahren besonders wichtige Veränderungen, vor allem im Hinblick auf die Arbeit mit Metadaten? Was und warum wurde da gemacht? o Haben Sie noch Überlegungen oder Ziele, was die nächsten Schritte für die Arbeit mit PHAIDRA sein könnten? Was könnte in PHAIDRA und im Bereich der Arbeit mit Metadaten dort noch weiterentwickelt werden? Weshalb gerade das?
• In welchem technischen Kontext fand die Arbeit an und Entwicklung der unterschiedlichen Metadatenformate für PHAIDRA statt? o Wie unterscheidet sich der Kontext, in dem die Entscheidungen damals getroffen wurden vom heutigen Kontext? o Welchen Herausforderungen sind Sie und das Team damals begegnet?

- Würden Sie mir berichten, welche Metadaten-Schemata momentan in PHAIDRA verwendet werden?
    - Weshalb werden mehrere Schemata verwendet? Was ist das Spezielle an den unterschiedlichen Schemata?
    - Wofür werden diese Schemata verwendet?
    - Was hat zur Entscheidung für gerade diese Schemata geführt?
    - Was sind die Vorteile, die sich dadurch ergeben?
    - Wie hängen diese Schemata mit der Eingabemaske in PHAIDRA zusammen und wie werden sie daraus generiert?
    - Weshalb werden nicht immer alle Formate generiert?

Themenblock 2: Herausforderungen und Entwicklungen in der Arbeit mit den Metadaten-Schemata

- Gibt es Herausforderungen, die mit der Zeit in der Arbeit mit den unterschiedlichen Metadaten-Schemata in PHAIDRA aufgetreten sind? Welche sind das?
    - Welche Rolle haben dabei Veränderungen und Entwicklungen, wie neue, wichtige Metadaten-Kategorien gespielt, die ergänzt werden mussten?
    - Welche Rolle haben aktuelle Entwicklungen im Bereich der Metadaten gespielt? Wie etwa der Wunsch, die Metadaten in Linked Open Data umzuwandeln oder die Möglichkeit, Metadaten mehr Nutzer*innen zugänglich zu machen?
    - Hätte man diese Entwicklungen und die Notwendigkeiten, die sich dadurch ergeben, voraussehen können?

Themenblock 3: Interoperabilität, einheitliche Erschließung und Zugänglichkeit der Metadaten

- Basierend auf Ihren Erfahrungen mit den unterschiedlichen Metadatenformaten in PHAIDRA, wie wichtig ist es Interoperabilität mit anderen Metadatenformaten zu ermöglichen?
    - Warum ist das wichtig?
    - Wie wird das am besten gemacht?
    - Welche Rolle spielt das für die langfristige Arbeit mit den unterschiedlichen Formaten?
    - Welche bestehenden Metadaten-Standards werden dafür verwendet? Etwa der Community und anderer Institutionen? Wie etwa Dublin Core, das ja auch ein Metadatenformat in PHAIDRA ist?
    - Werden, um Interoperabilität zu erreichen, die selbstentwickelten Metadaten-Schemata anderen Institutionen zugänglich gemacht? Wenn ja, auf welche Art?
    - Werden die Metadaten-Schemat auch in einem Repositorium abgelegt oder Aggregatoren oder Harvesting-Protokolle eingesetzt? Wird auch eine Metasuche unterstützt?
    - Gibt es, um die Metadaten anderen Institutionen zugänglich zu machen, auch ein Mapping zu anderen Metadatenformaten? Wenn ja, zu welchen Metdatenformaten und warum zu genau diesen?
    - Wenn Sie auf die Arbeit mit PHAIDRA blicken, haben Sie Empfehlungen, die Sie im Hinblick auf die Interoperabilität von Metadaten gerne weitergeben würden?
        - Welche Vorteile würden sich dadurch ergeben?
- Wenn Sie die Erschließung der digitalen Objekte durch unterschiedliche Nutzer*innen-Gruppen in PHAIDRA betrachten, wie werden die digitalen Objekte beschrieben?

- Warum ist das wichtig und welche Vorteile ergeben sich dadurch?
- Welche Rolle spielt das für die langfristige Arbeit mit Metadaten(-Formaten)?
- Wie wichtig ist dabei eine detaillierte Beschreibung der digitalen Objekte durch die Nutzer*innen?
- Welche Rolle spielen Pflichtfelder und optionale Felder? Wie werden diese festgelegt?
- Werden die von den Nutzer*innen vergebenen Metadaten auch außerhalb der Universität Wien zur Verfügung gestellt und weiterverwendet? Wie wird das gemacht?
- Gibt es Erkenntnisse aus Ihren Erfahrungen, die Sie gerne an die Verantwortlichen für zukünftige Repositorien weitergeben würden?
    - Welche Vorteile würden sich dadurch ergeben?

- Auf Basis Ihrer Erfahrungen mit PHAIDRA, werden Metadaten der digitalen Objekte in einem Repositorium für die interessierte Öffentlichkeit zugänglich gemacht?
    - Was spricht dafür, was dagegen?
    - Welche Ansätze sollten dafür verwendet werden?
    - Haben Sie in diesem Bereich Überlegungen, die für neu entstehende Repositorien wichtig wären?
        - Welche Vorteile würden sich dadurch ergeben?

Themenblock 4: Normierung und Dokumentation

- Wenn Sie auf Ihre Arbeit in PHAIDRA blicken, wie wichtig ist es, die Beschreibung der digitalen Objekte zu normieren und zu dokumentieren?
    - Welchen Zweck erfüllen die Normierung und die Dokumentation der Beschreibung?

- Welche Rolle spielt das für die langfristige Arbeit mit Metadaten?
- Wie werden die Normierung und Dokumentation der Beschreibung gemacht?
- Wie werden Schlagwörter aus standardisierten Vokabularen, Thesauren und Taxonomien verwendet?
- Wie sieht eine Dokumentation der verwendeten Vokabulare, Thesauren und Taxonomien aus?
- Wie werden die unterschiedlichen Bedürfnisse und Kenntnisse der Nutzer*innen im Hinblick auf die Verwendung dieser Standards berücksichtigt und ihre Erschließungsarbeit „angeglichen“?
- Gibt es Erkenntnisse, die Sie aus der Arbeit mit PHAIDRA gewonnen haben, die Sie gerne für die Arbeit an zukünftigen Repositorien weitergeben würden?
  - Welche Vorteile würden sich dadurch ergeben?

- Wie wichtig ist es, auch die verwendeten Metadaten-Schemata gut zu dokumentieren? Welche Vorteile bringt eine gute Dokumentation der Metadaten-Schemata?
  - Wie wichtig ist das für die langfristige Arbeit mit den unterschiedlichen Metadaten-Schemata?
  - Für wen ist diese Dokumentation zugänglich?
  - Welche Aspekte sind dabei besonders wichtig und werden festgehalten?
  - Gibt es darüber hinaus noch Informationen, die festgehalten werden?
    - Etwa zur erstellenden Institution, den verwendeten Standards, Thesauri und kontrollierten Vokabularen und dem Format, in dem die Metadaten dargestellt wurden?
    - Oder auch dazu, welche Anforderungen an Qualität und Vollständigkeit der Metadaten gestellt werden?
  - Gibt es Erkenntnisse aus dieser Arbeit, die Sie gerne für die Entwicklung neu entstehender Repositorien weitergeben möchten?
    - Welche Vorteile würden sich dadurch ergeben?

Themenblock 5: Langfristiger Erhalt und Pflege der Objekte durch Metadaten

- Basierend auf Ihren Erfahrungen mit dem Metadatenformaten in PHAIDRA, können Sie berichten, wie die langfristige Pflege und der Erhalt der digitalen Objekte durch Metadaten sichergestellt werden?
    - Welchen Zweck erfüllen diese Maßnahmen im Gesamten und welche Vorteile ergeben sich dadurch?
    - Welche Rolle spielt das für die langfristige Arbeit mit Metadaten?
    - Welche Aspekte und Metadatenkategorien werden dabei berücksichtig?
        - Gibt es beispielsweise Metadatenkategorien zur Entstehung, Zugriffsrechten, Inhalten, Weiterverwendung und eventuellen Verarbeitung der digitalen Objekte. Welcher Art sind diese?
        - Gibt es Metadatenkategorien, die etwa Format und Dateigröße der digitalen Objekte festhalten? Welcher Art sind diese?
        - Gibt es Metadatenkategorien, die Beziehungen zwischen den digitalen Objekten, ihrem Kontext, Verantwortlichen, Versionen, bisheriger Verarbeitung und Speicherung festhalten? Welcher Art sind diese?
        - Gibt es Metadatenkategorien, die festhalten, wann welche Objekte erstellt, verwendet und gepflegt wurden, und an die Institution weitergegeben wurden (für Statistik, Überwachung der Prozesse zur Berichtslegung administrativer, finanzieller und rechtlicher Aspekte)? Welcher Art sind diese?
        - Gibt es beispielsweise Metadatenkategorien, die es ermöglichen, digitale Objekte miteinander in Beziehung zu setzen, wenn etwa ein Objekt aus mehreren Seiten besteht (wie etwa ein digitalisiertes Buch)? Welcher Art sind diese?

- Was würden Sie davon halten, auch für die Metadaten einen persistenten Identifikator, wie etwa einen DOI oder einen Handle, einzusetzen? Weshalb?
- Gibt es Erkenntnisse, die Sie aus der Arbeit mit PHAIDRA in diesem Bereich gewonnen haben, die Sie gerne für die Arbeit an zukünftigen Repositorien weitergeben würden?
    - Welche Vorteile würden sich dadurch ergeben?

Themenblock 6: Festlegung der rechtlichen Nutzungsbedingungen

- Auf Ihrer Erfahrung mit den Metadaten-Schemata in PHAIDRA beruhend, gibt es eine Beschränkung der Nutzung der digitalen Objekte in digitalen Repositorien und wie wird diese in den Metadaten festgehalten?
    - Weshalb ist das wichtig?
    - Welche Vorteile bieten diese Metadaten zu den rechtlichen Aspekten für die langfristige Arbeit mit Metadaten?
    - Welche rechtlichen Aspekte sind dabei besonders zu berücksichtigen?
    - Werden urheberrechtliche Fragen berücksichtigt? Auf welche Art?
    - Wie wird die Wieder- und Weiterverwendung der digitalen Objekte geregelt?
    - Wie werden die Autorisierungen von Nutzer*innen, der Zugang zu den digitalen Objekten und die Überwachung der Zugriffe auf die Objekte verwaltet und kontrolliert?
    - Werden diese rechtlichen Aspekte in den Metadaten festgehalten und für Maschinen lesbar gemacht?Wie wird das gemacht?
    - Gibt es Erkenntnisse, die Sie aus der Arbeit mit PHAIDRA in diesem Bereich gewonnen haben, die Sie gerne für die Arbeit an zukünftigen Repositorien weitergeben würden?

▪ Welche Vorteile würden sich dadurch ergeben?

Themenblock 7: Zukunft der Metadatenformate in PHAIDRA
• Welche Umstände haben schließlich dazu geführt, verschiedene Mappings der UWMETADATA durchzuführen? o Welche Mappings wurden dafür angedacht und ausprobiert? o Was waren die Ziele dieser Mappings? o Welche Aspekte und Entwicklungen mussten dabei beachtet und miteinbezogen werden? o Welche Vorteile bieten diese Metadaten zu den rechtlichen Aspekten für die langfristige Arbeit mit Metadaten? o Gibt es Erkenntnisse, die Sie aus der Arbeit mit PHAIDRA in diesem Bereich gewonnen haben, die Sie gerne für die Arbeit an zukünftigen Repositorien weitergeben würden?
• Was sind die zukünftigen Pläne in Bezug auf die Metadaten und Metadatenformate in PHAIDRA? o Welche Ziele werden damit verfolgt?
• Was sind die nächsten Schritte, die in diesem Prozess angegangen werden? o Welche Herausforderungen gibt es noch dabei?

Schluss:

Vielen Dank für die Zeit, die Sie sich genommen haben und Ihre Teilnahme an diesem Interview.

Es freut mich sehr, dass Sie bereit waren, Ihr Wissen rund um diese unterschiedlichen Aspekte von Metadaten in PHAIDRA mit mir zu teilen.

Anhang 2: Interviewtranskripte

Anhang 2.1. Transkript des Interviews mit Raman Ganguly

Interviewpartner: Raman Ganguly

Datum: 20.12.2023

Ort: Wirtschaftsuniversität Wien

I:

Lieber Herr Ganguly, vielen Dank, dass Sie sich Zeit für das Interview nehmen. Es freut mich sehr. Und bevor wir beginnen, würde ich Ihnen zur Einführung gerne ein paar Fragen stellen. Und zwar würde ich Sie gern zuerst bitten, sich vorzustellen. Wie heißen Sie, wo arbeiten Sie und was genau ist Ihr Aufgabenbereich. #00:00:36#

B:

Okay. Mein Name ist Raman Ganguly. Ich arbeite an der Universität Wien, und zwar bin ich dort im Computercenter tätig, also im Zentralen Informatikdienst, so wie es offiziell heißt, und dort bin ich der Leiter der Abteilung IT-Support for Research. Das heißt, mein Hauptaufgabengebiet ist es, bei der Unterstützung von der Forschung das Thema Datenmanagement, und zwar die technische Infrastruktur dafür, zu übernehmen, beziehungsweise mir Konzepte und Infrastrukturen zu überlegen, wie man Datenmanagement machen kann und auch, was dafür notwendig ist, im Betrieb und in der Entwicklung. #00:01:24#

I:

Und spielen da digitale Repositorien und Metadaten auch eine Rolle? #00:01:30#

B:

Also digitale Repositorien sind einer von den zentralen Punkten. Es ist auch etwas, mit dem es mehr oder weniger begonnen hat, also den Repositories. Das, was wir heute Forschungsdaten-Repository nennen, das klassische, kommt ursprünglich aus dem Publikationsbereich her, das sind die Publikationsrepositorien gewesen. Und da orientieren wir uns sehr stark auch an den bibliothekarischen Metadaten, an den unterschiedlichen Standards, die es da gibt. Und so sind sie ursprünglich aufgebaut worden. Die Entwicklung ist jetzt ein bisschen woanders hin, aber ich denke, das ist im Interview dann auch, was noch besprochen wird. #00:02:10#

I:

PHAIDRA ist ja inzwischen schon 15 Jahre alt und Sie sind auch schon sehr lange beim ZID. Wie hat sich die Arbeit mit den digitalen Objekten und Metadaten, ihrer Vergabe und Darstellung in dieser Zeit entwickelt? #00:02:31#

B:

Also ich muss dazu sagen, ich war nicht von Anfang an bei PHAIDRA dabei. Also ich bin jetzt nicht die 15 Jahre bei PHAIDRA, bin zwar 15 Jahre am ZID, war vorher für etwas anderes zuständig, bin seit 2011 bei PHAIDRA, auch schon eine längere Zeit. (…) Was sich sehr stark verändert hat, ist der Fokus. Also zu Beginn war es ein sehr bibliothekarisches Thema, das ganze Thema Forschungsdaten-Repository, beziehungsweise Repository für digitale Objekte, und die Beschreibung dann auf einer bibliothekarischen Ebene. Es entwickelt sich mehr jetzt in diese Richtung, die Bedürfnisse von der Forschung und wie sie ihre Objekte beschreiben her zu betrachten. Und das ist nicht immer die gleiche Betrachtungsweise wie die bibliothekarische Betrachtungsweise. #00:03:31#

I:

Und das scheint so der wichtigste Prozess in den letzten Jahren gewesen zu sein. Also, wenn ich es richtig verstehe, PHAIDRA so zu gestalten, dass es möglichst gut die Forschungsdaten aufnehmen und darstellen kann. Gibt es irgendwas, was Sie sich wünschen würden für PHAIDRA, zukunftsorientiert, was noch möglich gemacht werden sollte, wohin es sich entwickeln sollte. #00:04:05#

B:

Also wohin es sich entwickeln sollte, das sind / Da gibt es, es gibt mehrere Ebenen. Ja, es gibt einmal die Ebene der Infrastruktur selbst, das Repositorium, das ja eine Eigenentwicklung ist von der Universität Wien, baut zwar auf Fedora auf, dem Repository, aber es ist relativ viel von uns auch noch entwickelt worden. Das eine Ding ist, hier mehr Nachhaltigkeit in der Entwicklung mit hineinzubringen, das heißt, mehr Richtung Open Source, Open-Source-Projektentwicklung zu gehen, das ist die eine Richtung. #00:04:44#

Die zweite Richtung ist dann eher der inhaltliche Teil: zu sehen, für was ist jetzt PHAIDRA, für welche Art von Daten ist PHAIDRA jetzt eigentlich das brauchbare Tool und dort hinzufokussieren. Und der dritte Bereich ist dann eher in dieses Thema FAIR-Data mit hinein, und zwar wirklich das Thema Findability, wo dann auch die Metadaten sich niederschlagen. Aber Findability jetzt auch im Sinne von dem her, wer sucht danach? Primär wären es im / oder wir wollen natürlich die Zielgruppe der Forscher*innen mehr ansprechen. Und hier sind es dann auch sehr domainspezifische Abfragen. Das heißt, wir müssen wahrscheinlich, oder das ist der Trend, den ich sehe, mehr Flexibilität auch in der Beschreibung der Metadaten oder der Objekte halt zu haben, im Sinne, die sich dann in den Metadaten niederschlägt, um die Findability zu erhöhen. #00:05:45#

I:

Da würde ich gerne kurz noch nachfragen, wie würde das aussehen, wäre das eine andere Verwendung von kontrollierten Vokabularen oder was würde da für unterschiedliche Communities den Zugang erleichtern? #00:06:02#

B:

Also, das eine wäre natürlich ein Idealszenario, das aber sehr schwer umzusetzen ist, weil auch eine zeitliche Thematik mit drinnen ist. Das ist einerseits, dass die Objekte mehrfach beschrieben werden. Das heißt, von unterschiedlichen Communities unterschiedlich beschrieben werden. Das ist insofern auch interessant für Objekte, die dann in unterschiedlichen Kontexten auch verwendet werden. Zum Beispiel gibt es ja gerade in der Lebenswissenschaft sehr viel interdisziplinäre Forschung. Was jetzt bedeutet, dass zum Beispiel Forschungsergebnisse aus der Chemie auch für die Biologie interessant sind, umgekehrt

auch wieder für die Biologen, die Chemie. Und dann diese Forschungsbereiche, die sowieso beides inkludieren müssen, wie zum Beispiel die Pharmazie, die einerseits von der Biologie Daten hat und andererseits auch von der Chemie die Daten hat. Die Schwierigkeit in dem Bereich ist es, dass Chemiker die Dinge anders beschreiben als Biologen und umgekehrt. Und wenn jetzt diese Objekte, die relevant sind, aus mehreren Ebenen beschrieben werden, dann wären sie natürlich für mehrere Communities leichter zu finden, weil einfach gewisse Terminologien verwendet werden. #00:07:18#

Eine Hilfe ist natürlich, Terminologien mehr und intensiver einzusetzen. Hier haben wir aber natürlich auch wieder dieses Thema mit drinnen, dass ein gewisser Aufwand mit dabei ist. Ich denke, dass in Zukunft, und das wird jetzt der spannende Bereich, auch für die Repositorien-Welt und für die Datenwelt sein, hier den Einsatz von KI zu forcieren, nämlich das automatisierte Beschreiben von diesen Objekten, und das aus unterschiedlichen Perspektiven zu sehen. Also das sehe ich auch als Zukunftstrend, noch nicht aktuell, weil momentan ist es ein großer Hype und man muss schauen, was da übrigbleibt und nicht bei jedem Hype halt dann mitspringen, sondern wirklich schauen, wo ist dann auch die Qualität dahinter? #00:08:05#

I:

Wir bewegen uns damit quasi in einem ganz neuen, relativ neuen technischen Kontext, der sich wahrscheinlich enorm davon unterscheidet von dem Kontext, der bestand als PHAIDRA und die Metadatenformate entwickelt worden sind. Was war damals anders? #00:08:30#

B:

Ich denke, damals war es einfach (…) die Sichtweise eine ganz eine andere. Also, wir haben 2007, also „wir“ inkludier ich jetzt auch mich, obwohl ich erst 2011 dazugekommen bin, aber von den Ideen her / auf diesen Ideen, die entwickelt worden sind, wurde dann aufgebaut. Also 2007 ist die Idee entstanden, und sie ist aus einem bibliothekarischen Kontext heraus entstanden. #00:08:59#

Es ist in der Bibliothek dieses Projekt entwickelt worden. Mit der Idee, ich habe digitale Objekte, also ich habe im Endeffekt Assets, digitale Assets und die möchte ich im Endeffekt so beschreiben, wie ich auch ein Buch beschreiben möchte, damit ich die wiederfinden kann wie ein Buch. Und auch in diesem Sinn eine langfristige Aufbewahrung habe, wie bei einem Buch. Also das waren die zwei Zielsetzungen, die es damals gab und aus einer sehr traditionellen Sichtweise, was jetzt die Wissensaufbewahrung betrifft, gesehen worden sind. #00:09:37#

Hier ist natürlich mit dem, dass jetzt die (…) Personen mit hineingekommen sind, die auch die Daten aktiv entwickeln, das heißt, die auch mit die Daten dann produzieren, aber sie auch nachnutzen, beziehungsweise, wir wollen ja, dass sie nachgenutzt werden, haben wir hier gesehen, dass es hier ein /, dass wir aus unserer Sichtweise heraus uns bewegen müssen oder ändern müssen, weil es ist doch eine sehr geschlossene Sichtweise. Und diese Geschlossenheit muss aufgebrochen werden. Das sehen wir nicht nur jetzt im Sinn von einem Repository, sondern das sehen wir in sehr vielen anderen Bereichen auch, dass dieses Aufbrechen gerade stattfindet, und das ist ein genereller, ich will jetzt nicht sagen Trend, aber es ist eine gewisse Notwendigkeit, die die Digitalisierung mit sich bringt, ja. #00:10:38#

I:

Also genau das, was Sie vorhin erwähnten, die Objekte und die Metadaten unterschiedlichen Communitys zur Verfügung zu stellen und auch deren Standards oder Wünsche zu berücksichtigen. #00:10:55#

B:

Es geht hier jetzt nicht um Standards primär. Die Standards, denke ich, werden sich im Laufe der Zeit entwickeln. Das heißt, auch im Bereich der Standards ist eine Bewegung mit drinnen. Es geht mehr in diese Richtung / Ich würde jetzt nicht sagen / Sie haben es mit Bedürfnis oder Wünsche bezeichnet. Ich würde es eher da aus einer technischeren Sichtweise betrachten, die Use Cases einfach. Ich möchte ja, dass das ganze Ding nutzbar wird und nutzbar wird es dann, wenn die Leute es in ihrer Arbeitswelt integrieren können. Ansonsten ist es eine Pflicht, die getan werden muss, weil es in irgendeiner Art und Weise

vorgeschrieben wird. Das soll es ja nicht sein. Es soll nutzbar sein und da muss ich mich dann auf ihre Arbeitsweise, beziehungsweise auf ihre Kultur, auch Arbeitskultur einstellen. Und da sollte ein Repository auch mehr oder weniger kein Störfaktor sein. #00:12:02#

I:

Ich würde Sie gern viel mehr noch dazu fragen, aber ich kehr mal zurück zu den Metadaten in PHAIDRA. Können Sie mir berichten, welche Schemata momentan verwendet werden, weshalb mehrere verwendet werden und warum gerade diese? #00:12:24#

B:

Also, da bin ich wahrscheinlich nicht die beste Quelle. Also Frau Blumesberger weiß davon mehr, weil sie A seit dem Beginn mit dabei ist und B sich auf dieses Thema Metadaten spezialisiert hat. (…) Soweit ich das kenne oder weiß, wie dieses Uni-Wien-Metadaten-Schema entstanden ist, war es eine Zweckgemeinschaft von verschiedenen Standards, sagen wir mal so. #00:12:55#

Wir hatten ursprünglich auch neben dem, dass die Bibliothek mit dabei war und der ZID als die beiden Treiber, dieses Repository zu entwickeln, auch das Center for Teaching and Learning mit dabei. Das ist bei uns die E-Learning-Einheit der Universität, und dadurch waren hier auch diese Bedürfnisse berücksichtigt. Das heißt, es waren nicht nur die bibliothekarischen Bedürfnisse berücksichtigt, die in Richtung, der Beschreibungen, was da aus bibliothekarischer Sicht gekommen ist, dann ist auch das da, was aus einer Lehr-Sicht gekommen ist. Das heißt, dieser LOM-Standard, und dann war natürlich auch das Bedürfnis aus einer technischen Sichtweise, dass wir ja mit diesem Repository die Metadaten austauschen wollten mit anderen, das heißt, dass darum auch das Dublin Core ganz stark mit drinnen ist und wir am Anfang auch sehr viele Wünsche berücksichtigt haben, und dadurch ist es zu einem sehr großen (…), nicht einem Standard entsprechenden Metadatenkonglomerat gekommen, das wir dann UW-Metadaten genannt haben. #00:14:15#

Von da haben wir dann verschiedene Standards herausdestilliert. Eines der wichtigsten ist natürlich UWMETADATA / ist natürlich das Dublin Core. Wir haben dann auch bei

einigen Projekten jetzt das MODS mit hineingegeben, wo wir ein Mapping auch gemacht haben. Und wir gehen auch im Bereich von / Also das E-Learning ist dann aus dem Projekt ausgeschieden, aber nach etlichen Jahren durch ein anderes Projekt wieder dazugekommen. Und wir haben hier jetzt dann auch wieder das LOM-Schema mehr oder weniger in Verbindung. Das heißt, wir haben immer gewisse Ausschnitte. Was zwar in der Handhabung sehr schwierig ist, was uns aber jetzt in der Änderung der verschiedensten Szenarien, wo wir das Ganze einsetzen wollen, beziehungsweise wo wir überall auch die Metadaten hinschicken möchten, uns etwas leichter tun. #00:15:17#

I:

Ja, weil es ganz viele Ansatzpunkte gibt, ja. ,00:15:20#

B:

Genau, ja. #00:15:21#

I:

Ich habe mir ein paar digitale Objekte und die Metadaten dazu angeschaut. Die Metadaten, wie werden die aus der Eingabemaske generiert? Also, und es werden ja auch nicht immer alle generiert. #00:15:49#

B:

Wir haben gewisse Pflichtfelder. Es gibt ein paar, die automatisch reinkommen, Datum-Sachen und solche Sachen. Dann gibt es Pflichtfelder. Und das war es. #00:16:01#

I:

Okay. #00:16:02#

B:

Das ist das, was ist, was wir jetzt über diese Eingabemaske haben. Dann gibt es aber auch spezialisierte Eingabemasken, die jetzt für gewisse Projekte auch stattfinden, die können andere Pflichtfelder definieren. Dadurch haben sie eine andere Metadatenqualität. Für die Masse müssen wir die Stufe sehr einfach machen, dort hineinzubringen. Wir haben aber

auch dann die Möglichkeit, dass wir über eine Schnittstelle Daten hineinbringen, und dort gibt es verschiedene Metadaten, also Metadaten-Quellen, die wir dann auf die Objekte mappen. Je nachdem, wie gut die Quelle ist, umso besser sind die Metadaten. #00:16:42#

I:

Gehen wir zum nächsten Themenblock. Gab es Herausforderungen, die mit der Zeit in der Arbeit mit den unterschiedlichen Metadaten-Schemata in PHAIDRA aufgetreten sind? (…) Gab es zum Beispiel wichtige neue Metadaten-Kategorien oder (…) eher technische Entwicklungen oder Wünsche, die Metadaten in Linked Open Data umzuwandeln. Oder wie Sie vorhin auch erwähnt haben, die Metadaten anderen Nutzer*innen zugänglich zu machen. #00:17:18#

B:

Also, ich würde sagen, diese ganzen Thema Bedürfnisse, Metadaten, neue Metadaten oder sonst irgendwas, da, da kann ich wenig dazu sagen, das wäre nett (…) / kann ich nicht gut beantworten. Das Thema Linked Open Data ist natürlich ein sehr großes Thema, aber das ist eher auch ein Thema, das die technische Infrastruktur betrifft. Also wir können im Endeffekt, da haben wir die Metadaten derzeit in einem XML-Format vorliegen. Mit einer neuen Version von Fedora gehen wir sehr stark in diesen Triplestore hinein. Das heißt, das ist im Prinzip dann Linked Data. Das sind einfach technische Themen, die aufgrund von Änderungen der Infrastruktur auch notwendig werden. Also das sind so Migrationsthemen, in denen sind wir ganz stark natürlich mit drinnen, hat aber jetzt aus meiner Sicht, oder aus einer sehr naiven Infrastrukturbetreibersicht, wenig mit dem zu tun, wie sie genutzt werden. Ja, also da würde ich eher dann Herrn Hudak oder die Frau Blumesberger diesbezüglich mehr fragen. #00:18:26#

Was ich herausforderungsmäßig / da kann auch dann der Herr Hudak besser dazu antworten, ist es / Wenn wir natürlich verschiedene Metadatenschemen mit drinnen haben: Man muss sich auf einen gewissen Core einigen, der dann in die Suche mit hineingeht, weil sonst hat man diese generische Suche über alle Objekte, da tut man sich sehr schwer, was dann gefunden wird und was nicht gefunden wird, das hat auch hier Implikationen in die Usability dann wieder hinein. #00:18:53#

I:

Also in diesem ganzen PHAIDRA-Kontext und der Arbeit mit PHAIDRA und Metadaten: Wie wichtig ist es, Interoperabilität von Metadatenformaten zu ermöglichen? Warum und wie wird es am besten gemacht? #00:19:16#

B:

Interoperabilität ist einmal eins von diesen wichtigen Merkmalen, wenn ich jetzt in Richtung Forschungsdatenmanagement denke, ist es das, das das Paradigma ist: Daten müssen den FAIR-Prinzipien entsprechen. Und das I steht für Interoperabilität, das heißt, das ist ein Musskriterium, damit wir als Infrastruktur überhaupt sinnvoll mit dabei sein können. Also es ist eine zwingende Notwendigkeit. Also von daher ist es ein ganz wichtiger Punkt. #00:19.58#

Das Thema bei der Interoperabilität ist natürlich das: Ich kann immer sagen, okay, ich nehme das und das Schema, aber auch innerhalb von einem Schema gibt es Interpretationsspielraum und das haben wir ganz gut gesehen, wie wir mit Europeana zusammengearbeitet haben, dass hier teilweise die Dinge unterschiedlich interpretiert worden sind und es auch nicht so trivial war diese beiden Dinge zusammenzubringen. Das heißt, Interoperabilität ist ein großes Ziel. Ich sehe es aber relativ schwer. (…) rein aus einer Sicht, aus einer Infrastruktur-Sicht quasi zu sehen, wenn ich jetzt die isoliert betrachte. Das heißt, ich muss auch das Gegenüber betrachten. Das heißt, mit wem möchte ich interoperabel sein. Und da sind natürlich dann diese großen (…) Trends dann, wo wir hingehen müssen, also European Open Science Cloud oder andere große Forschungsinfrastrukturen. Und an denen werden wir uns in Richtung Interoperabilität orientieren müssen. In diese Richtung geht es. Also, von dem her ist es eigentlich eins von den Dingen, die unsere Pflichtübung betreffen. #00:21:17#

I:

Sie sprechen von der Infrastruktur. (…) Was also müsste auf formaler Ebene / Müssten die Schemata wahrscheinlich miteinander interoperabel sein? Aber zusätzlich müsste man sich auch also zwischen den Institutionen entscheiden, wie man das angehen möchte? Also gibt, gibt es da unterschiedliche technische Ansätze, wie das möglich wäre?

B:

Also, das Thema Interoperabilität ist wieder die Frage: Mit wem möchte ich interoperabel sein? Also, das ist genau das Thema. Und ja, da muss man sich im Prinzip auf organisatorischer Ebene mal klar werden und auf Dinge einigen. Das ist einmal ganz wichtig (…) und dann muss es dazu technische Lösungen geben. Die technischen Lösungen sind eher nicht so, das Triviale/ ist eher nicht so das Schwierige, ja. Das, was nicht trivial ist, ist auf organisatorischer Ebene, sich auf etwas zu einigen. Ja, und das ist ganz wesentlich und ich denke, auch hier werden wir in den nächsten Jahren sehr starke Änderungen haben, weil wir mehr intelligentere Schnittstellen haben werden, die einfach Transformationen automatisiert machen können. Und auch hier, denke ich, ist der Bereich Machine Learning, also auch wieder in Richtung AI ein ganz wesentlicher Treiber, wo wir einiges verändern können, beziehungsweise erleichtern können. #00:23:00#

Es wird nicht so sein, dass wir / Also das Thema ist viel zu komplex, auf das, dass wir uns auf einer sehr simplen Ebene einigen können, wie zum Beispiel Steckersteckdosen. Also ich kann mit jedem Gerät, kann ich hergehen und mich egal in Österreich irgendwo anstecken und es funktioniert ja. Wenn ich in die USA gehe, haben wir schon wieder ein Problem. Das heißt, auch hier gibt es im Prinzip Unterschiede. Das heißt, auch das ist nicht interoperabel. Ich brauche ein Zwischenstück, einen Adapter dafür, ja. Und diese Adapter, je komplexer das wird, umso komplexer oder beziehungsweise, umso häufiger werden diese Adapter auch werden, und das wird ein großes Thema in der Zukunft werden, damit ich diese Interoperabilität damit leisten kann. Was ich eher hoffe, dass nicht passiert, ist, dass die Vielfalt, die zurzeit da ist, dass die mit diesem großen Thema Interoperabilität sinken wird, weil dann haben wir wieder nur ein paar große Player, die bestimmen, was darf und was nicht sein darf. #00:24:04#

I:

Wie viel haben Sie mit der Erschließung der digitalen Objekte durch die Nutzer*innen zu tun? #00:24:22#

B:

Gar nichts. #00:24:22#

I:

Gar nichts. #00:24:23#

I:

Dann noch eine Frage zu den Metadaten der digitalen Objekte. Werden die irgendwo extra abgelegt, in einem eigenen Repositorium, oder sind die nur über PHAIDRA zu finden? #00:24:40#

B:

Also die Metadaten per se sind nur über PHAIDRA zu finden. Also (…), das ist auch wieder so ein Thema. Die Objekte werden einmal in PHAIDRA aufgenommen. Kann aber sein, dass es ein korrespondierendes (…) Katalogverzeichnis gibt. Also beispielsweise ist es, wenn ich jetzt hergehe und ich digitalisiere ein Buch von der Bibliothek, dann sind die Metadaten bereits erfasst im Bibliothekskatalog. Das heißt, ich erfasse sie nicht noch einmal neu, das wäre sinnlos, beziehungsweise eine unnütze Arbeit. Das heißt, hier ist der Workflow dann so geregelt, dass die Objekte mit den Metadaten aus dem Bibliothekskatalog gematcht werden und die dann in PHAIDRA aufgenommen werden. Bei diesen Workflows muss man immer definieren, wer ist das datenführende System. In dem Fall ist es der Bibliothekskatalog, weil dort der Katalog ist oder dort dieses Ding ist, wo die höchste Expertise ist. Das heißt, Bibliothekare arbeiten eher mit dem Bibliothekskatalog als wie mit PHAIDRA. Kann aber natürlich sein, dass es bei anderen Dingen anders ist. #00:25:57#

Oder, wie ich schon gesagt habe, es gibt Automatismen, wo ich Daten übernehme, wo die Metadaten irgendwo anders bereits liegen. Das heißt, auch hier liegen sie dann parallel wo. Also, das heißt, das kann man global so nicht sagen, per se ist es so, wir sehen die Metadaten und Objekt als eine Einheit. Ich kann dann von dieser Einheit, kann ich natürlich Teile davon mit anderen teilen, das heißt, die Metadaten per se sind immer offen und ich kann meine Metadaten mit anderen Katalogen teilen, dass diese Information dann auch woanders ist. Ja, also das, das ist ein Standard, der in PHAIDRA drin ist, beziehungsweise ein Standard eines jeden Repositories, dass ich meine Metadaten teilen kann. #00:26:41#

I:

Und Aggregatoren und Harvesting-Protokolle sind auch einfach Standard?#00:26:50#

B:

Genau. Ja, ja. Also OAI-PMH und wir können auch über unsere API gehen und können hier auch die Metadaten abholen. Also es gibt verschiedenste Möglichkeiten, wie wir das machen. Also das Teilen der Daten und das ist jetzt wieder, auch wenn ich auf die FAIR-Data Prinzipien zurückkomme, das wäre das Reusable, das heißt, das ist genauso ein wichtiger Punkt wie das Interoperable und / auf diese Ebene hin designen wir uns auch unsere Infrastruktur. Das heißt, wir ermöglichen es, einen leichten Zugang zu den Metadaten und zu den Objekten zu haben. Wobei die Objekte wieder ein bisschen anders sind, weil es ja auch gesperrte Objekte sein können, und dort soll der Zugang natürlich nicht so leicht sein. #00:27:40#

I:

Okay, gehen wir zur Normierung und Dokumentation über. Wie wird die Normierung und Dokumentation gemacht? Welchen Zweck erfüllt sie? #00:28:06#

B:

Der Objekte jetzt, oder? #00:28:06#

I:

Genau, der digitalen Objekte. Ich dachte an standardisierte Vokabulare, Thesauren, auch Taxonomien, wie die eingesetzt werden und wie sie dokumentiert werden und ob den Nutzer*innen diese Dokumentation zugänglich ist für die Beschreibung der digitalen Objekte, weil die machen sie ja selbst. #00:28:28#

B:

Genau, die Beschreibung machen sie selbst. Wir haben einige Thesauren bei uns mit integriert in PHAIDRA. Das kann ein jeder nutzen. Wir haben das im Endeffekt nicht großartig dokumentiert. Das ist die Frage, wie weit wir das, / also das ist auch eine Frage, die ist jetzt strategisch eher mehr an der Bibliothek angesiedelt, die sind nicht bei uns. Weil

wir sind nur die, die dann quasi die Infrastruktur zur Verfügung stellen. Was ist, wir haben eher den Ansatz der Schulungen und wir haben den Ansatz, mit dem, dass wir versuchen, möglichst früh mit den Forscher*innen in Kontakt zu treten und denen die Möglichkeiten zu zeigen. Die Dokumentation und das Handhaben mit den einzelnen Objekten überlassen wir denen, weil das ist ja ihre Aufgabe, und wir haben nicht die Ressourcen dafür, das zu machen. Das würde eher in die Data Curation mit hineinfallen, was wir nicht tun, also was wir nicht leisten können in einer Größenordnung von der Universität. #00:29:32#

Ob sich das in den nächsten Jahren ändert, beziehungsweise wie viel Ressourcen wir in dieses Thema mit hineinnehmen, kommt auch ein bisschen dazu von einerseits A, was wird gefordert, quasi von außen, Forschungsförderer und so weiter. Und B wie schaut die Ressourcenlage innerhalb von der Universität aus, und C wie sieht die Universität selbst das Thema Forschungsdatenmanagement. Weil bis jetzt haben wir als Infrastrukturprovider mehr oder weniger definiert, wie es genutzt werden soll. Was ein bisschen eigenwillig ist oder was im Prinzip nicht / Ich meine, was immer der Beginn von etwas Neuem ist, ist, dass die, die das Neue zur Verfügung stellen, auch definieren, wie wird es genutzt. Erst ab dem Zeitpunkt, wo die Nutzung halt einen gewissen Schwellenwert überschreitet, muss es eine übergeordnete Governance auch geben, die sich darum kümmert, wie dann das genutzt werden soll. #00:30:30#

Ja, also ganz praktisch, die ersten, die ein Kraftwerk gebaut haben und den Strom zur Verfügung gestellt haben und die sich dran angehängt haben, die Haushalte, ich weiß nicht, vor 150 Jahren oder so, irgendwas in dieser Größenordnung, da hat es noch keine Regelungen gegeben. Die Regelungen sind ja später geworden, wie dann mehr oder weniger, das weiter ausgebaut worden ist. Und jetzt gibt es sehr starke Normierungs-Regelungen, jetzt gibt es E-Control und so weiter und so fort. Also auch hier wird sich etwas tun. Momentan sind wir noch in einem etwas freieren Umfeld. Das heißt, wir geben auch sehr viel Freiheit, die wir haben, die geben wir weiter. Wird sich aber sicher in den nächsten Jahren massiv ändern. In welche Richtung kann ich noch nicht sagen. #00:31:14#

I:

Also, vielleicht werden die Vorgaben auch ein bisschen strenger, oder? #00:31:20#

B:

Die Vorgaben werden ganz sicher strenger werden. Es ist immer so, dass man von sehr frei zu sehr streng wechselt. Dann merkt man, dass sehr streng auch nicht funktioniert, dann öffnet man das ganze wieder. Also das heißt, wir werden immer diese Pendelbewegung haben, und mal schauen, wo sie sich dann mehr oder weniger einpendelt. Ich denke, wir haben schon die Tendenzen in Richtung strenger zu werden. Also allein was Forschungsförderer betrifft, an Regulierungen beziehungsweise an Vorgaben. Data-Management-Plan ist zum Beispiel ein Thema, das wäre im Sinne der Dokumentation auch so die Geschichte. Aber das obliegt wieder den Forschungsprojekten. #00:32.02#

I:

Und gibt es solche Vorgaben auch für administrative Objekte, die ja zum Teil in PHAIDRA abgelegt /? # 00:32:12#

B:

Für welche Projekte? #00:32:13#

I:

Administrative Objekte? #00:32:16“

B:

Wir haben keinen Fokus jetzt auf administrative Objekte momentan. Kann sich aber ändern, in dem Sinn, dass wir auch mit dem Universitätsarchiv gerade im Gespräch sind, wie sie PHAIDRA nutzen können. Hier würde aber auch dann, also wir würden hier wieder in diese Rolle gehen, dass es eine Ownership der Daten gibt und die, die den Ownership haben, definieren auch die Regelungen. Auf der infrastrukturellen Ebene werden wir so wenig Regelungen wie möglich einführen, damit wir auch offen sein können für Neuentwicklungen. #00:32:53‘

I:

Gilt das auch für die Dokumentation der Metadaten-Schemata oder haben Sie da für sich, oder das Team, hat das für sich klare Vorgaben, dort wird dokumentiert, dort wird abgelegt. #00:33:12#

B:

Also wie gesagt wieder zurückkommen auf das, was wir eigentlich tun. Wir tun, wir machen die Technik, wir bauen die Kraftwerke. Und auf dieser Ebene haben wir natürlich alles dokumentiert. Wir haben Informationssysteme, wir wissen, wo der Source Code liegt, haben den Source Code dokumentiert. Auf dieser Ebene ist alles klar, wie das dann genutzt wird, das obliegt wieder den Nutzern. Also da wieder eher dann mit der Bibliothek reden. #00:33:38#

I:

Ja. Ich frag einfach nach. (…) Das Gleiche gilt wahrscheinlich auch für den langfristigen Erhalt und die Pflege der Objekte durch Metadaten. Da hat auch die Bibliothek einfach die Expertise. #00:33:56#

B:

Genau. Also, es gibt gewisse Objekte, die dann in das eher, mehr oder weniger in den Ownership der Bibliothek wechseln, wie zum Beispiel die Open-Access-Publikationen. Dafür sind sie auch für die Metadaten zuständig, das heißt, es obliegt im Prinzip den Ownern der Objekte, wie sie damit umgehen. Es wird sicher einen sehr großen Bereich geben, wo die Metadaten einmal erfasst werden und dann nie wieder angeschaut. Es wird aber sehr viel anderes auch geben. Das heißt, das delegieren wir zu denen, die das dann im Endeffekt nutzen, ja. #00:34:31#

I:

Okay. (…) Gilt das auch für die rechtlichen Nutzungsbedingungen und Metadaten, ist dafür auch die Bibliothek ausschließlich zuständig? #00:34:42#

B:

Da muss man jetzt ein bisschen unterscheiden. (…) Wir haben den Ansatz, dass wir infrastrukturell alle Objekte aufnehmen, die irgendwo auf der Universität halt dann entstehen. Jeder kann es nutzen. Also wir sind völlig offen in der Nutzung. Wie die Objekte genutzt werden, obliegt denen, denen die Objekte gehören. Das heißt, die müssen auch gewisse Regelungen halt dann einsetzen. Das heißt, auch die Bibliothek ist auf unterschiedlichen Ebenen, einerseits Serviceprovider, andererseits Kunde von PHAIDRA, ja. #00:35:26#

Die Abteilung von der Susanne Blumesberger ist die, die das Service PHAIDRA zur Verfügung stellt. Dann gibt es aber zum Beispiel das Open-Access-Büro der Bibliothek, die Nutzer von PHAIDRA sind, weil die Open-Access-Dokumente drinnen sind. Aus Sicht der Bibliothek als Nutzer definieren sie, wie mit den Metadaten umgegangen wird und wie sie / und sie sind dann auch zuständig für die Pflege der Metadaten und für eine langfristige Verfügbarhaltung der Metadaten, also aus dieser Sicht und auch für das ganze rechtliche Thema. #00:36:00#

Auf der Infrastruktur-Ebene, das heißt, wie wir dieses Gefäß PHAIDRA, wie das genutzt werden darf, da gibt es auch Nutzungsbedingungen. Diese Nutzungsbedingungen definiert im Prinzip die Service-Abteilung PHAIDRA, aber das natürlich in Absprache oder in gemeinsamer Zusammenarbeit mit uns. #00:36:22#

Das heißt, man muss hier immer die Ebenen unterscheiden und die Ebenen sind sehr unterschiedlich. Das heißt, von dieser Ebene, ja. Wir sind zum Beispiel auch zuständig für die langfristige Aufbewahrung der Objekte. Das heißt, wir schauen auch aus technischer Sicht, dass sie mehr oder weniger aufbewahrt werden können. Wir geben Empfehlungen oder nehmen nur gewisse File-Formate an. Beziehungsweise sagen halt dann, diese File-Formate sind langfristig aufbewahrungswürdig und diese nicht. #00:36:57#

Aber dann, wenn es in Richtung Migration von diesen Daten geht, dann muss man sich dann eh noch was überlegen, wie man die Qualitätssicherung machen kann, weil die Qualitätssicherung kann man nicht ohne die ursprünglichen Owner der Objekte machen, weil wir kennen diese Objekte ja nicht. Von daher muss man, muss man immer diese Ebenen unterscheiden und auch alles, was die rechtliche Nutzung und so irgendwas von

PHAIDRA betrifft, das obliegt uns. Das heißt, wenn wir zum Beispiel mitbekommen, dass rechtswidrige Materialien in PHAIDRA sind, sind wir auch dafür zuständig, dass die gelöscht werden. Die Klärung der Urheberrechte von einem Objekt reinzulegen, obliegt wieder dem, der dieses Objekt im Prinzip besitzt, also dem Owner. #00:37:49#

I:

Ja, und welche möglichen Arten von Rechten vergeben werden können, dafür ist die Bibliothek verantwortlich? #00:37:55#

B:

Nein, der, der das Objekt hineinlädt. #00:37:58#

I:

Also er ist da völlig frei.

I:

Mhm. Also wir sind sehr offen. Also das ist das, was wir, was wir gesagt haben, das liegt wahrscheinlich auch an dem, dass wir schon so früh angefangen haben mit dem Thema, dass wir da noch die Möglichkeit gehabt haben, wo es keine Regulierung gegeben hat, oder weniger Regulierung gegeben hat, dass wir sehr offen damit umgehen können. Und wir bauen dann eher in diese Richtung auf, dass wir sagen, okay, wir bauen, wir haben die Infrastruktur und wir haben Services. Diese beiden trennen wir aber sehr stark. Das heißt, wir etablieren gerade eine Rechtsberatung für Forscher*innen, damit die im Prinzip Rechtssicherheit haben, wie sie dann die Lizenzen innerhalb von PHAIDRA vergeben können oder dürfen. Das heißt, es wird aber auf einer anderen Stelle geklärt, das wird nicht bei uns geklärt in der Infrastruktur. #00:39:05#

I:

Kommen wir zur Zukunft der Metadatenformate in PHAIDRA, wenn Sie dazu noch was sagen möchten. Welche Umstände haben denn dazu geführt, dass mehrere Mappings von UWMETADATA durchgeführt werden? Haben Sie da Kontaktpunkte gehabt? #00:39:27#

B:

Na ja, das große Bedürfnis war einfach das: Es hat verschiedene Anwendungsfälle gegeben, wo es notwendig war und das wird auch weiter so sein. Also wann es neue Anwendungsfälle gibt, und wir können diese bedienen, wird das hier Änderungen geben. Und ich bin mir ziemlich sicher, dass es hier massive Änderungen noch geben wird. Und ich bin mir auch ziemlich sicher, dass wir an irgendeinem Punkt Konsolidierungen machen müssen, weil es nicht mehr möglich ist, diese Vielfalt zu haben. Die Frage ist es, können wir es abstrahieren, technisch abstrahieren, und an dem bin ich auch schon länger am Überlegen, wie weit kann ich wirklich die Datenstruktur und Metadaten-Schema hier abstrahieren, damit es mir aus einer Infrastrukturebene wieder völlig egal ist, was für ein Schema da eingesetzt wird. Das ist ein Wunsch, den hätte ich, aber der ist sehr schwierig zu realisieren, beziehungsweise zurzeit sehe ich noch keine vernünftige Lösung. #00:40:34#

I:

Das wäre dann auch ein Wunsch für die Zukunft. #00:40:39#

B:

Das wäre ein Wunsch für die Zukunft, weil dann habe ich als Infrastruktur-Provider überhaupt kein Thema mehr. Dann soll auch jeder diesen Metadatenstandard verwenden, den der Owner von den Objekten als notwendig empfindet, genauso wie mit den Lizenzen umzugehen. #00:41:03#

I:

Gibt es Pläne für die Zukunft? Also, das ist der Plan, für das ist der Wunsch für die Zukunft, der Plan für die Zukunft sich dem anzunähern? #00:41:15:#

B:

Also es gibt, es gibt momentan viel dringlichere Dinge, also wie dieses Thema mit den Metadaten. Wieder auf die Infrastruktur runtergebrochen ist das Allerwichtigste für mich momentan aus strategischer Sicht her, was die Infrastruktur betrifft, ist die nachhaltige

Sicherung der Infrastruktur als solches. Weil wenn ich hergehe und ich sage, ich habe ein Repository für die Langzeitarchivierung, aber das Repository selbst ist nicht nachhaltig gesichert, dann habe ich ein bisschen Argumentationsschwierigkeiten. Und das ist der Punkt, wo ich mich jetzt mehr oder weniger darauf konzentriere, wie kann ich es schaffen, eine nachhaltige Lösung für die Weiterentwicklung dieses Repositories zu generieren. #00:42:04#

I:

Ja, wir sind dann schon am Ende angekommen. Vielen herzlichen Dank, dass Sie da waren und dass Sie Ihr Wissen mit mir geteilt haben, das war ein sehr spannendes Gespräch. Danke. #00:42:18#

B:

Freut mich, vielen Dank. #00:42:19#

Anhang 2.2. Transkript des Interviews mit Susanne Blumesberger

Interviewpartnerin: Susanne Blumesberger

Datum: 15.01.2024

Ort: Wirtschaftsuniversität Wien

I:

Liebe Frau Blumesberger, herzlichen Dank, dass Sie sich heute Zeit nehmen für dieses Interview. Zunächst möchte ich Sie bitten, sich kurz vorzustellen. Wie heißen Sie, wo arbeiten Sie und was genau ist Ihr Aufgabenbereich? #00:00:18#

B:

Sehr gerne. Vielen Dank für das Interview. Es bringt mich selber vielleicht auf neue Gedanken. Ich freue mich, dass Sie mich eingeladen haben. Danke schön. #00:00:24#

Mein Name ist Susanne Blumesberger und ich arbeite an der Universität Wien in der Bibliothek, in der Universitätsbibliothek, seit 2007. Und habe dort mit dem Aufbau auch begonnen des institutionellen Repositoriums PHAIDRA. #00:00:39#

Und meine derzeitige Aufgabe ist die Leitung der Abteilung Repositorien Management PHAIDRA Services, wie wir es genannt haben, jetzt. Weil PHAIDRA inzwischen nicht nur ein System ist, sondern mehrere Komponenten hat und dazwischen, also dazu auch noch alle möglichen anderen bibliothekarischen Aufgaben. Wie zum Beispiel auch internationale Projekte, nationale Projekte, also Vernetzungen und so weiter, Nachwuchsförderung, ich unterrichte auch und so. #00:01:06#

I:

In diesem komplexen Aufgabengebiet, das Sie gerade beschrieben haben, wie wichtig ist da die Arbeit mit digitalen Repositorien, PHAIDRA selbstverständlich, aber auch mit Metadaten? #00:01:20#

B:

Ja, selbstverständlich ist das ein wichtiger Punkt bei uns. Also unser Repositorium wurde ja 2008 sozusagen in die Welt gesetzt. Seitdem gibt es das Repositorium. Und die Arbeit an den Metadaten, also mal zu überlegen, welche Metadaten man überhaupt benötigt in einem Repositorium, das war ein ganz ein wichtiger Punkt und hat eigentlich das erste Jahr richtig ausgefüllt. Also wir haben 2007 begonnen. Da startete das Projekt, so hieß es damals: Projekt Digitale Langzeitarchivierung. Und wir haben wirklich ein Jahr lang versucht herauszufinden, welche Metadaten von welchen Personen gebraucht werden. #00:01:56#

Also wir sind dann wirklich in der Uni Wien herumgegangen, haben uns Pilotpartner gesucht. Aus unterschiedlichen Bereichen, zum Beispiel aus der Translationswissenschaft, aus der Bibliothek, aber auch aus der Physik, anderen Bereichen, und haben einfach geschaut: Was brauchen die Wissenschaftler? Was brauchen die Forschenden, um eben ihre Objekte gut beschreiben zu können? #00:02:17#

Das war ein ganz ein wichtiger Punkt, weil natürlich, wie man sich vorstellen kann, die Uni Wien ist riesig und sehr, sehr heterogen. Das heißt, wir haben unglaublich unterschiedliche Daten und die alle unter einen Hut zu bringen und zu sagen, ich kann die jetzt einfach in einem Repositorium so beschreiben, dass man wirklich das Gefühl hat, man hat die Daten gut beschrieben und das kommt eigentlich dem zugute, was man wirklich auch möchte. Das heißt, man trifft wirklich genau den Punkt, ist gar nicht so einfach, und das war für uns eine ganz wichtige Phase natürlich. #00:02:48#

I:

Es ist dann wahrscheinlich auch nicht einfach, allen (…) Stakeholdern gerecht zu werden, also den unterschiedlichen Fachgebieten mit den unterschiedlichen / #00:03:02#

B:

Absolut, absolut, das schaffen wir auch sicher überhaupt nicht, weil für die einen ist es zu viel Information und für die anderen ist es viel zu wenig. Oder die falsche. Also zum Beispiel, ich kann mich erinnern, am Anfang war das Institut für Kunstgeschichte sehr interessiert an PHAIDRA. Die haben schon richtig gewartet, dass sie endlich so eine Möglichkeit bekommen, ihre Daten langfristig zu sichern. Und die haben sich ganz genau

angeschaut, welche Metadaten da möglich sind. Die haben zum Beispiel Dinge gebraucht wie Stempel. Das war so ein typischer Fall, also etwas, was sonst niemand wirklich damals gesagt hat, aber sie mussten einfach die Stempel, die zum Beispiel auf ihren Glasplatten-Dias drauf waren, die wollten sie dort beschrieben haben, und solche Dinge kamen. #00:03:42#

Also abgesehen davon, dass wir natürlich von Anfang an mehrsprachig unterwegs waren. Wir haben auf jeden Fall Englisch natürlich angeboten, dann aber auch Italienisch, weil wir sehr schnell italienische Partner hatten. Also das war wirklich von Anfang an ein ganz ein wichtiger Punkt für Metadaten. #00:03:58#

I:

Das war vor etwa 15 Jahren. Also PHAIDRA ist jetzt schon 15 Jahre alt. Und wie hat sich dann die Arbeit seit damals weiterentwickelt? Einerseits mit dem Repositorium, aber auch mit den Metadaten, wie sie vergeben werden und dargestellt werden, gab es da unterschiedliche Phasen, wichtige Veränderungen? #00:04:27#

B:

Also die Arbeit hat sich auf jeden Fall verändert, weil natürlich wir / Also sie verändert sich jederzeit, weil wir natürlich unterschiedliche Personen haben, die etwas in PHAIDRA stellen. #00:04:42#

Es ist ja bei uns so, dass jede Person an der Uni Wien, auch die Studierenden, berechtigt sind, etwas in PHAIDRA zu stellen, und zu sagen, ich möchte das langzeitarchivieren. Und das machen nicht wir von der Bibliothek, sondern das machen die Personen selbst. Das heißt, die haben die Berechtigungen mit ihrem Account, mit dem u:account einzusteigen und einfach loszulegen, ja. #00:04:58#

Und das hat / Das verändert sich natürlich je nach Projekt. Also wir haben inzwischen unterschiedlichste Möglichkeiten, die wir anbieten können. Also ausgegangen sind wir eigentlich von einem Einzel-Upload der Projekte, ah, der Objekte. Das hat sich dann rasch herausgestellt, dass das nicht ausreicht, weil dann hat man zum Beispiel Bildersammlungen, und da will man nicht 1.000 Fotos einzeln hochladen, klarerweise. Das heißt, wir

brauchten einen Bulk-Upload. Da wurden dann unterschiedliche Masken erstellt vom Zentralen Informatikdienst, mit dem wir sehr eng zusammenarbeiten, also der das mit aufgebaut hat. Und die haben dann Masken erstellt und es bestehen immer noch Masken für unterschiedliche Projekte. Und dann gibt es ständig eine Weiterentwicklung, ein Weiterdenken, ein Umdenken also, das reißt auch nicht ab. #00:05:45#

I:

Ja, und in diesem Kontext verändern sich dann auch immer die Metadaten, weil wenn ich mir dann zum Beispiel vorstelle, so ein Bulk-Upload, da werden wahrscheinlich auch die Verbindungen der Objekte untereinander beschrieben, wie die zusammenhängen, die Objekte selbst, wahrscheinlich bei Bildern Größe, Format und solche Sachen. #00:06:04#

B:

So ist es. Bei Bildern /…/ Also wir haben einen großen Sprung, muss ich sagen, haben wir voriges Jahr gehabt. Wir hatten sehr lange das Gleiche, die gleiche Maske sozusagen, wo man hochlädt. Und voriges Jahr war es dann endlich so weit, dass wir eine neue Maske hatten. Das heißt, wir haben jetzt eine Maske, wo wir je nach Objekttyp, also ob ich jetzt ein Video hochlade, oder einen Text oder ein Bild, je nachdem habe ich andere Metadatenfelder zur Verfügung. Was auch sinnvoll ist, weil die Leute dann einfach schneller was einspielen können und nicht mehr durch alles durchgehen müssen, was sie gar nicht betrifft. Das war vorher der Fall. Ja, also da hat sich schon einmal einiges geändert. #00:06:41#

Das geht jetzt viel rascher und wir haben natürlich auch Dinge dazubekommen. Ja, also zum Beispiel bei Projekten Projektnummern und so wurden dann nach und nach noch dazugenommen. Dinge, die vorher vielleicht gar nicht so relevant sind, wo wir draufgekommen sind, ja, das sollten wir eigentlich auch noch mit abdecken. Dann die Pflicht-Metadatenfelder, die haben sich verändert. Also wir hatten am Anfang, glaub ich, sechs oder sieben, dann waren es nur mehr fünf und jetzt sind es acht, weil wir draufgekommen sind, wir brauchen doch ein bisschen mehr Informationen von den Leuten. Und es ist gut, das zu haben. Also für die Objekte, Wiederauffindbarkeit. #00:07:16#

I:

Ok, weil sie /, weil dadurch die Objekte leichter identifiziert werden können, durch / #00:07:22#

B:

Genau, zum Beispiel bei Projektnummer ist es ganz wichtig, die zu haben. Das war nicht von Anfang an drinnen. Die Zugehörigkeit, das ist auch ein Punkt zum Beispiel. Also wie gesagt, wenn jemand was hochlädt, dann lädt er das ja unter einem bestimmten, von einem bestimmten Institut oder von einem Department oder so hoch. Und das wussten wir früher nicht. Da wurde halt hochgeladen und jetzt haben wir das verpflichtend gemacht. Das heißt, man muss angeben, ich arbeite da jetzt für das Institut für Biochemie und dann gebe ich das an, ja. Also das ist dazugekommen. Die ORCID-ID zum Beispiel ist auch etwas, was man jetzt eintragen kann, was in Zukunft noch viel stärker / eine stärkere Rolle spielen wird. #00:07:59#

I:

Sehr gut. Das ist eine (…) sehr spannende Entwicklung und, dass ORCID (…)/ also, die ORCID verwenden zu können. Das wird einfach auch immer wichtiger, denke ich. #00:08:16#

Können Sie sich erinnern, in welchem technischen Kontext die Arbeit, als PHAIDRA startete, stattgefunden hat, in Bezug auf die Metadaten, und wie unterscheidet sich das vom heutigen Kontext? Also, Sie haben es eh ein bisschen beschrieben. (…) Aber gab es, ja Sachen, die Sie damals gerne gewusst hätten? #00:08:50#

B:

Ja, gibt es schon (lacht). Na, also, PHAIDRA ist ja eigentlich eine / basiert auf einer Open Source Software Fedora. Und der ZID, der Zentrale Informatikdienst bei uns, hat das weiterentwickelt. Und wir haben von Anfang an einige Metadaten-Schemata drinnen gehabt in PHAIDRA, Dublin Core zum Beispiel, und das LOM-Schema – Learning Object Metadata. Das hatten wir deshalb drinnen, weil wir auch die Lehre sehr dabeihaben wollten. Und das hat sich jetzt als sehr günstig erwiesen, weil inzwischen gibt es, also eh

schon länger, Projekte zu Open Educational Resources, die genau dieses Metadaten-schema benötigen, um eben diese Objekte gut beschreiben zu können. #00:09:36#

Und als wir dann begonnen haben, DOIs zu vergeben, das hatten wir von Anfang an nicht. Wir hatten die PHAIDRA-Links und wir haben dann irgendwann begonnen, Handle-Links zu vergeben. Und das ist jetzt nach wie vor so, dass jedes Objekt einen Handle-Link bekommt und eine DOI auf Anfrage. Ja und DOI haben wir über DataCite gemacht. Das heißt, wir haben auch DataCite jetzt drinnen. Das hatten wir auch nicht von Anfang. #00:09:59#

I:

Aber sonst ist noch UWMETADATA /? #00:10:01#

B:

Genau, das ist so ein bisschen selbstgestrickt, kann man sagen. Das sind eben all diese Dinge, die nirgends reingepasst haben, die aber manche unserer User, Userinnen gebraucht haben. Die haben wir einfach dazu mit hineingenommen. #00:10:14#

I:

Das sind dann einfach Kategorien, also das sind zum Beispiel Metadaten-Kategorien, die die User*innen zusätzlich gebraucht haben. #00:10:00:20#

B:

Die sie zusätzlich gebraucht haben, die aber, zum Beispiel diese Stempelgeschichte, die es sonst nicht gibt. Das zum Beispiel, das ist sehr spezifisch in PHAIDRA. #00:10:30#

I:

Und JSON-LD ist jetzt ein neues /, oder? 00:10:39

B:

Das ist relativ neu. Ja, das kommt auch daher, dass wir wirklich / also auch die Technik im Hintergrund verändert oder upgedatet wurde. Also wir sind sehr, sehr lange auf dieser Fedora 3.8 gewesen. Und jetzt geht es in Richtung 6. Es ist gleich ein großer Sprung, weil

es irgendwie einfacher ist, von dem / Ich kenne mich technisch nicht aus, aber die Techniker sagen mit es immer; von dem ist es besser wegzuspringen, sozusagen als von der Vierer-Version, oder von der Fünfer. Und das hat sich auch damit entwickelt, dass man auch dieses Triple, RDF hineingebracht hat, zum Beispiel. Also da gibt es wirklich viele Änderungen inzwischen, die da im Hintergrund sind, die aber den Usern jetzt gar nicht so auffallen und das ist uns wichtig, die sollen ja davon gar nichts mitbekommen. #00:11:25#

I:

Wissen Sie zufällig / also diese unterschiedlichen Metadatenformate / Sie haben gesagt, es gibt ein paar Pflichtfelder. Und wenn ich es richtig verstehe, werden dann aus der Eingabemaske die Metadatenformate generiert? Wovon hängt das ab, welche generiert werden? Oder werden immer alle generiert? #00:11:47#

B:

Soweit ich weiß, werden immer alle generiert. Also was eingegeben wird, also diese Pflichtfelder, die kommen zum großen Teil aus Dublin Core. Also so wie Autor, zum Beispiel, oder dass man ein Abstract eingeben muss, Schlagworte vergeben. Was haben wir dann noch? Ja, den Autor eben mit der Personen-ID, wie auch immer, oder ohne, geht auch. Das kommt alles aus dem Dublin Core in Wirklichkeit. #00:12:16#

I:

Wir haben schon ein bisschen darüber gesprochen / die Herausforderungen, die dann aufgetreten sind mit der Zeit. Neue Metadaten-Kategorien mussten ergänzt werden. (…) Und eine Frage hätte ich dazu noch. (…) Hat auch die Idee jetzt Richtung Linked Open Data zu gehen und diese Darstellung, in Richtung RDF zu gehen, hat das auch eine Rolle gespielt? Spielt es momentan eine Rolle bei der Überlegung, wie mit den Metadaten umgegangen wird? #00:13:03#

B:

Ja, also da sind wir / Also da möchten wir auf jeden Fall hin. Auf Linked Open Data sind wir noch nicht wirklich dort, aber es wird versucht. Es wird zum Beispiel bei den ganzen

Thesauri versucht, da ein bisschen in die Richtung zu gehen. Da geht auf jeden Fall der Trend hin. Also wir versuchen immer mehr zu standardisieren. Also das war am Anfang vielleicht nicht ganz so, und das ist jetzt schon ziemlich stark, also wir versuchen wirklich, möglichst standardisiertes Vokabular zu verwenden, zum Teil, was aber nicht wirklich durchhaltbar ist, weil das teilweise nicht passt. #00:13:34#

Also ich habe gestern wieder / Wir haben wieder ein Gespräch gehabt mit Kolleginnen und Kollegen von der Kultur- und Sozialanthropologie. Und wir haben ja ÖFOS, also von der Statistik Austria, verpflichtend. Es ist eines der Pflichtfelder, die man eingeben muss. Und sie haben gesagt, sie fühlen sich da überhaupt nicht vertreten. Das passt einfach nicht. Und das haben wir schon öfter gehört, ja. #00:13:55#

Oder manche sagen, sie wollen unbedingt die gemeinsame Normdatei, die GND drinnen. Da haben wir inzwischen auch die Möglichkeit, da einzutragen und auch da sagen die Leute, das passt für uns nicht, das ist nicht ausreichend, das ist nicht granular genug. #00:14:10#

Dann teilweise gibt es das nicht auf Englisch, was auch ein Problem ist. Es hakt noch. Also das ist ganz, ganz / Das ist wirklich etwas, was ganz wichtig geworden ist. Das Englische, das spüren wir auch so. Aber auch bei den Metadaten, dass natürlich die Leute sehr viel auf Englisch eingeben, also je nach Fachgebiet, also die arbeiten alles auf Englisch ab. #00:14:32#

I:

Ich frag Sie / Wir sind schon bei den Thesauri und den Vokabularen, deswegen würde ich Sie dazu gerne noch etwas fragen. Also diese, diese standardisierten Thesauri und Vokabulare, die ermöglichen ja einerseits so eine (…) gewisse Vergleichbarkeit, werden, wie Sie aber gerade gesagt haben, nicht den Ansprüchen der Nutzer*innen gerecht, nicht in allen Fällen. (…) Arbeiten Sie da mit international verbreiteten / also die GND, ja selbstverständlich im deutschsprachigen Raum, ÖFOS haben Sie genannt / Sie arbeiten eben auch immer mit standardisierten Vokabularen und Thesauri und (…) die werden auch dokumentiert? #00:15:52#

B:

Ja, die werden dokumentiert und es gibt auch die Möglichkeit, selbst Vokabularien einzuspielen. Das heißt, wir haben einen sogenannten Vocabulary Server und da kann man, wenn man das möchte und Vokabularien schon hat, für irgendeine Fachrichtung sagen, ich möchte genau das verwenden. Die können dann das einspielen sozusagen und verwenden. Das geht auch. #00:16:16#

I:

Und, also parallel zu der Dokumentation der Thesauri und Vokabulare, werden die Metadaten-Schemata auch dokumentiert? Und wem ist die Dokumentation dann zugänglich?

B:

Mit dokumentiert ist gemeint, dass es dahintersteht, wie das ausschaut, also was reinkommen soll, also Beispiele oder so? #00:16.37#

I:

Also ja, zum Beispiel Beispiele, aber auch die Definition der einzelnen Metadatenfelder. Also was soll da rein und dann vielleicht ein Beispiel, in welcher Reihenfolge sie auftreten müssen. Gibt es sowas? #00:16:56#

B:

Also es gibt auf jeden Fall so Hilfstexte, die man bekommt, wenn man was einfügt. #00:17:02#

I:

Ja, die tauchen dann so als Mouseover auf? #00:17:05#

B:

Also, genau, die tauchen dann als Mouseover auf. Dann gibt es auf der Serviceseite Informationen über die Metadaten. Aber so richtig, dass man jetzt sagt, genau, also wir haben schon natürlich verlinkt auf die Dublin Core-Seite zum Beispiel. Das schon, aber / also darüber hinaus wüsste ich jetzt nicht, was man den Leuten angeben kann, weil es ist

für manche sehr schwierig zu verstehen, was sie überhaupt dort einfügen müssen. #00:17:34#

I:

Und haben Sie eine interne Dokumentation für das Team? #00:17:38#

B:

Ja, wir haben schon natürlich, wir haben natürlich schon intern / Was Sie vorhin gesagt haben, mit Mapping, haben wir auch gemacht und versucht, da möglichst konsequent zu sein: Bei den Rollen zum Beispiel, das ist ein großes Thema, Rollen, ja. Man glaubt nicht, wie lange man über Rollen diskutieren kann. In welcher Rolle lad ich jetzt hoch? Also da gibt es ja die Fotografin, den Fotografen, den Uploader, den Autor, die Autorinnen, dann gibt es den Interview-Partner. Aber dann gibt's ja noch ganz für uns abstruse Bezeichnungen, wie Kupferstecher oder so. Ist eh klar, braucht man natürlich auch. Und das hat sich in letzter Zeit ziemlich ausgeweitet, je nachdem welches Projekt wir gerade sozusagen servicieren. #00:18:18#

Wir sagen dann oft nur eigentlich, diese Rolle brauche ich unbedingt, dann gibt es diese Rolle vielleicht nicht in irgendeinem Vokabular im MARC oder so. Dann müssen wir etwas, ja, dazunehmen. Dann gibt es noch die Kategorie Anderes, das kann man dann noch dazunehmen, weil das vielleicht nur eine Gruppe von Personen von sieben Leuten in ganz Österreich braucht. Ja, so ungefähr (lacht). Aber es ist wichtig für die, sie möchten das einfach haben und sie haben auch einen Grund, warum sie das möchten, weil sie es eben abbilden müssen. Also da haben wir schon eine Dokumentation für uns intern, genau. #00:18:50#

I:

Gibt es in dem Bereich also /.../ Thesauri, standardisierte Vokabulare, Dokumentation von Metadaten / haben Sie da Erfahrungen gesammelt, wo Sie denken, das wäre gut gewesen, wenn wir das gewusst hätten, und das könnten wir jetzt auch an andere Repositorien-Betreiber weitergeben? #00:19:15#

B:

Na ja, es sind andere / Wir sind ja inzwischen viele Partner, also wir haben über 22, glaub ich jetzt, fast schon 23 sind es inzwischen Partner, die auch PHAIDRA verwenden. Ja, wir haben das nicht angestrebt, aber das hat sich so entwickelt. Und denen geben wir selbstverständlich das weiter. Also das ist überhaupt so, dass alles, was getan wird, was entwickelt wird, ist Open Source und kann man sich in Wirklichkeit holen. Von daher, alles, was wir jetzt lernen, und was weiterentwickelt wird, steht den anderen zur Verfügung und das war einfach ein Prozess, wo man nicht sagen kann, ich hätte es gern vor zehn Jahren schon gewusst, weil jetzt brauchen wir es, jetzt haben wir es. Jetzt haben wir es entwickelt. Es hat sich nichts geändert. Natürlich ist das alles eine Frage der Entwicklung und selbstverständlich, ja, wenn man das vor zehn Jahren gewusst hätte, hätten wir es damals schon reingeben können, weil die Leute dann mehr Metadaten eingetragen hätten, aber das kann man schwer sagen. #00:20:07#

I:

Jetzt würde ich Sie gern noch was zur Interoperabilität fragen. (…) Wie wichtig ist es, Interoperabilität zu ermöglichen und warum ist es so wichtig? #00:20:25#

B:

Also das war ein wirklicher Punkt, den wir von Anfang an mitgeplant haben, dass wir eben kein Silo sind, wo die Leute was reinstellen. Und dann ist es da drinnen und dann können die nicht mehr irgendwo das weiterverbreiten. Sondern es war von Anfang an mitgedacht, dass wir ja, also einen Austausch haben wollen mit anderen Systemen. Und da waren wir recht früh auch mit OpenAIRE zum Beispiel in Kontakt. Also ich glaub seit 2009, dieses EU-Projekt, die auch harvesten, ja, die sich auch Daten holen und Europeana. Das war auch / vor allem Europeana Libraries, das war ein großes Projekt bei uns. Das uns sehr geholfen hat, weil da sehr viel digitalisiert werden konnte und andererseits natürlich auch genau dieser Austausch angeregt werden konnte. Ja, also das, das war von Anfang an ein großer Punkt. Das haben wir immer gesagt, dass das wichtig ist, ja. #00:21:15#

I:

Und da wurden einfach auch unterschiedliche technische Ansätze eingesetzt, um das zu ermöglichen. Ich habe schon ein Gespräch mit Herrn Ganguly gehabt. Und er hat mir erzählt, dass es gar nicht so leicht ist, unterschiedliche Metadaten-Schemata, jetzt auch zum Beispiel mit Europeana zusammenzubringen, weil auch immer jede Institution für sich, also das Schema selbst für sich definiert. / #00:21:50#

B:

Und interpretiert / Ja, ja, das stimmt, das stimmt ja, es ist gar nicht so einfach, wie man vermuten möchte. Also man kann auch Dinge wirklich falsch verstehen, oder falsch interpretieren, so wie „publisher“ zum Beispiel. Weil das mehrere Bedeutungen hat, dann weiß man oft nicht, was man reinschreiben soll. Dadurch, dass er bei uns ja nicht begleitet wird, dieser Prozess des Hochladens. Und die Leute einfach sitzen und sich denken, aha, das klingt nach dem, ich schreib das rein, haben wir das auch nicht wirklich in der Hand. Was die Leute reinschreiben. #00:22:22#

I:

Ja, die, also sie /…/ die Nutzer*innen, sind eigentlich bei der Beschreibung unabhängig. Und das, also, hat wahrscheinlich Vorteile für die Nutzer*innen, weil die einfach entsprechend ihren Bedürfnissen eintragen können, aber dann für die Verwendung der Daten und Metadaten, kann, wenn dann jemand Fachfremdes kommt oder es zu einem Austausch kommen soll / Wird es dann schwierig? #00:22:46#

B:

Könnte unter Umständen schwierig werden. Ja, also ich muss dazu sagen, weil sowohl OpenAIRE als auch Europeana nehmen ja nur basic Daten, also die nehmen ja wirklich nur ganz, ganz / die nehmen nur gerade den Titel, eine Beschreibung, Autor oder Datum. Mehr glaube ich / und die Lizenz natürlich. Da kann man jetzt nicht so viel falsch machen, sage ich einmal. Alles, wo es dann wirklich um Dinge geht, die man, ja, weiter interpretieren kann, wie zum Beispiel Herkunft des Objekts, Provenienz, so, da kann man irgendwas reinschreiben, das wird ja meistens gar nicht mitgeliefert. Das ist ja auch kein Problem, das passt dann für die Leute, die es hochgeladen haben, und für die Leute, die es abrufen, die haben dann halt eine zusätzliche Information. #00:23:29#

I:

Also Sie haben ja selbst schon gesagt, die Objekte in PHAIDRA, PHAIDRA ist kein Silo, und die Objekte sind allen zugänglich, also auch der interessierten Öffentlichkeit. Da kann jeder eigentlich schauen, was es gibt. Und auch durch die Kooperation mit Europeana wird das ja nochmal auf eine weitere Stufe gehoben. # 00:24:11#

B:

Genau. Genau. Also, es gibt die Möglichkeit, Objekte zu sperren. Ja, das müssen wir machen, weil natürlich viele Objekte nicht gleich zumindest offen sein können, aus unterschiedlichen Gründen. Das kann man granular sperren, also man kann es auf die Uni sperren, man kann es auf diverse Institute oder Departments sperren, man kann es auf Personengruppen, man kann es auf Einzelpersonen sperren. Und mit der Zeit noch kombinieren. Das heißt, ich kann auf fünf Jahre etwas sperren. Die Metadaten allerdings sind immer sichtbar. Die bleiben auch sichtbar, wenn das Objekt gesperrt ist. Das war uns ganz wichtig, damit eben die Forschung sichtbar wird. Auch wenn ich sage, ich kann das jetzt nicht freigeben, weil, keine Ahnung, ich möchte das publizieren, ich möchte das irgendwie noch nicht, dass es jemand weiß, was ich da gemacht habe, aber die Metadaten bleiben sichtbar. #00:24:57#

I:

Wir sind da schon bei letztendlich rechtlichen Fragestellungen, weil es / also einerseits geht es um die Forschung, an der noch gearbeitet wird, die noch nicht vollständig publiziert werden soll. (…) Und da gibt es dann rechtliche Beschränkungen durch diese Sperren, also die Zugänglichkeit. Welche rechtlichen Aspekte werden noch berücksichtigt bei der Arbeit mit diesen digitalen Objekten? Wie werden Rechte vergeben, welche werden vergeben? Also ich denke da an einerseits an so Zugriffsrechte, aber auch wie wird mit dem Urheberrecht umgegangen? Wie wird dann auch die Wieder- und Weiterverwendung, was ja ein ganz großes Thema ist in der Forschung, wie wird das geregelt? Zu dem Zugang für Nutzer*innen, zu den Metadaten haben sie Zugang, der Rest kann gesperrt werden? Wie arbeitet da PHAIDRA? #00:26:03#

B:

Also zunächst einmal, es gibt Nutzungsbedingungen, die akzeptiert werden müssen, bevor man etwas hochladen kann. Und dort steht drinnen, dass man eben nichts hochlädt, was irgendwelchen Rechten widerspricht. Das heißt, man sagt, man hat die Rechte auf dieses Objekt. Damit ist einmal die Person, die hochlädt, verantwortlich für das, was diese Person hochlädt. Ja, das ist einmal der Ausgangspunkt. #00:26:26#

Und dann habe ich zwei Möglichkeiten. Ich kann Objekte sperren, damit sind sie nicht sichtbar. Ich kann Objekte aufmachen und sie so lizenzieren, dass diese Objekte nur angesehen werden dürfen. Ich kann zum Beispiel sagen, ich vergebe hier keine Creative Commons Lizenz, sondern ich sage das Urheberrecht sozusagen, steht da drüber und man kann dieses Objekt zwar downloaden, man kann es sich anschauen, aber man darf nichts damit tun. Und durch diese Creative Commons Lizenzen, die wir verwenden, ist genau geregelt, was darf die Person, die das Objekt im Netz findet, tun? Oder jemand anderer. Und das heißt dann, ich kann zum Beispiel das völlig frei verwenden, ohne irgendetwas damit, also ohne irgendwie tun zu müssen. Oder ich darf es nur in der Form verwenden, wie ich es vorfinde, ich darf es nicht verändern, ich darf es nicht verkaufen und so weiter und man kann das auch kombinieren. #00:27:25#

I:

Ja genau, dieses non-commercial, no derivatives #00:27:29#

B:

Genau. Genau. #00:27:30#

I:

Und diese / Da habe ich eine kurze Frage: Und ist in PHAIDRA angelegt, dass vor allem diese Creative Commons Lizenzen ausgewählt werden können? In den Metadaten, quasi? Und die sind dann also, wenn sie dann in den unterschiedlichen Formaten dargestellt werden, sind sie auch maschinenlesbar letztendlich? #00:27:51#

B:

Genau. Also das ist der Punkt. Also man gibt das wirklich / beim Hochladen muss man sich entscheiden, entweder man gibt keine Lizenz, dann kann man es verändern. Wenn man einmal eine Lizenz gewählt hat, kann man diese nicht mehr verändern. Dann bleibt das so. #00:28:05#

I:

Das ist auch eine interessante Vorgabe. Also da darf man keinen Fehler machen. #00:28:10#

B:

Nein, das sollte man nicht und ich sage den Leuten immer, bevor sie es genau wissen, welche Lizenz sie brauchen, weil oft bei Forschungsprojekten, ja, was mache ich, was muss ich jetzt tun, keine Ahnung. Dann keine Lizenz vergeben, weil von der keinen Lizenz kann man immer noch auf eine Lizenz umschwenken und sagen, ich vergebe diese Lizenz. #00:28:25#

I:

Das ist gut zu wissen. Damit sind wir / haben wir noch ein Thema. Wir haben es ein bisschen angesprochen, die unterschiedlichen Metadaten, die für die langfristige Pflege und den Erhalt verwendet werden. #00:28:45#

Sie haben schon erzählt, also es gibt einerseits diese Metadaten, die beschreiben, wie etwas entstanden ist, also so administrative Metadaten, dann gerade bei Bilddateien die technischen Metadaten zu Größe und Format und die strukturellen Metadaten. Darüber haben wir auch schon gesprochen. Gibt es auch Metadaten (…), die (…) in PHAIDRA vergeben werden, die eher für eine Art Berichtslegung verwendet werden, innerhalb der Institutionen. Weil ich mir zum Beispiel vorstellen könnte, dass die Bibliothek ja irgendwie auch belegen muss, wie PHAIDRA genutzt wurde, was hochgeladen wurden, wie die Objekte weitergegeben, vielleicht auch weiterverwendet werden, gibt es /? #00:29:47#

B:

Ja, also die Zuordnung ist ein Punkt, ja, dass wir eben diese Zuordnung, die zu den jeweiligen Instituten drinnen haben, was Verpflichtung ist. Das haben wir noch nicht sehr lange, vielleicht zwei Jahre oder so. Damit man eben sagen kann, das Institut für, ich weiß nicht, Kunstgeschichte hat so viel drinnen, und das Institut für Germanistik hat so viel raufgeladen, also das ist für interne Zwecke. #00:30:06#

Was wir auch haben, ist eine Zugriffsstatistik. Also das wird auch mitgezählt, wie viel Zugriffe es auf die Objekte gibt. Wobei das halt immer nur auf dem Weg geht, und nicht auf anderen Wegen. Wenn das Objekt woanders gefunden wird, dann haben wir natürlich keinen Einfluss drauf. Und eine Downloadstatistik läuft auch im Hintergrund mit. Was haben wir sonst noch? Ja, ich glaub, das ist es eigentlich, was wir das so im Hintergrund haben. #00:30:36#

Also was vielleicht auch noch spannend bei PHAIDRA ist, weil das hat, glaub ich, sonst kein Repositorium: Wir ermöglichen den Personen auch, die Objekte zu beschreiben, die analog vorhanden sind, das heißt, wenn ich jetzt irgendein / Nehmen wir ein Bild, ein gemaltes Bild. Ich kann dieses Bild fotografieren und kann dieses Foto digitalisieren und in PHAIDRA stellen. Und dann sage ich, ich habe es upgeloadet an dem Tag und dem Tag, keine Ahnung. Ich kann aber auch dieses upgeloadete Foto dann so beschreiben, dass ich dieses Bild dahinter beschreibe. Das heißt, ich kann sagen, der Maler so und so hat das gemalt, 1789 oder ich weiß nicht, sowas in diese Richtung, ja. Also diese Provenienz, glaub ich, das glaub ich hat sonst kaum ein Repositorium, was auch ermöglicht, wirklich, den Sammlungen gerecht zu werden. Ja, den Sammlungsdaten auch, ja. #00:31:30#

I:

Ja, mit einer beschränkten Anzahl an Metadatenfeldern trotzdem allen gerecht werden. #00:31:36#

B:

Ja, genau. Also das, das führt auch immer wieder zu / nicht zur Verwirrung. Aber gestern hatten wir auch mit der KSA, mit der Kultur- und Sozialanthropologie ein Gespräch, wo es darum ging, das Objekt auf dem Bild zu beschreiben. Das heißt, nehmen wir an, ich

habe, ich habe ein Glasplatten-Dia und auf dem Glasplatten-Dia, gestern haben wir das Beispiel gehabt, ist ein Webstuhl abgebildet. Ja, so, jetzt kann ich das digitalisierte Objekt beschreiben. Ich kann das Glasplatten-Dia beschreiben, weil auf dem Glasplatten-Dia, ich weiß nicht, ob Sie das kennen, das ist größer und da steht manchmal irgendwas drauf, da stehen noch ein paar Begriffe drauf. Die kann man auch mit hineinnehmen. Und dann kann ich aber auch den Webstuhl beschreiben. Und dann machen sich mehrere Ebenen auf, plötzlich. Wie bilde ich das ab, ohne dass es zu komplex wird, dass es überhaupt gar nicht mehr verstanden wird, ja? #00:32:26#

I:

Eine schwierige und spannende Frage. Eine Frage hätte ich da noch. Sie haben vorhin den DOI und den Handle erwähnt, die Objekte kriegen einen Handle und auf Anfrage einen DOI. Kriegen die Metadaten auch einen Handle und einen DOI? #00:32:43#

B:

Nein, die Metadaten selbst nicht, weil die sind veränderbar. Die Metadaten können die Owner jederzeit verändern. Das wird zwar mitdokumentiert, dass es verändert wurde. Aber man kann, weiß nicht, wenn ich jetzt draufkomme, ich habe vor fünf Jahren was reingestellt in PHAIDRA und habe dort einen Tippfehler oder ich habe eine zusätzliche Information, die ich unterbringen möchte, kann ich das jederzeit machen, also. Und dadurch, dass das verändert wird, wird das nicht viel Sinn machen. Und der DOI, der soll ja ohnehin nicht auf das Objekt selbst weisen, sondern auf eine Landing-Page. Und diese Landing-Page enthält ja die Metadaten, das heißt ich komme ja ohnehin auf die Metadaten. #00:33:22#

I:

Okay, dann wäre das eigentlich gar nicht notwendig. #00:33:27#

B:

Nein, meiner Meinung nach nicht. #00:33:29#

I:

Dann zum letzten Themenblock. (…) Welche Umstände haben denn schließlich dazu geführt, dass verschiedene Mappings der UWMETADATA durchgeführt wurden? #00:33:46#

B:

Das waren zum Teil wirklich auch die Anforderungen der, erstens der Partneruniversitäten, die vielleicht anderes gebraucht haben, und natürlich auch der Projekte, wo wir mitgemacht haben, also OpenAIRE und Europeana zum Beispiel. Kulturpool ist noch eines, wo es geheißen hat, wir brauchen unbedingt, keine Ahnung, das und das von euch, weil sonst können wir das nicht . #00:34:12#

I:

Ich glaub MODS war/ (…) und dann wahrscheinlich JSON-LD, das Neueste. Und das hat dann einfach die Zusammenarbeit erleichtert zwischen den Institutionen? #00:34:28#

B:

Den Austausch, genau, das hat einfach den Austausch dann erleichtert. Und man muss natürlich trotzdem immer schauen, weil beim Kulturpool ist es zum Beispiel jetzt so, da haben wir schon, ich weiß nicht, wie lange das Projekt jetzt läuft, aber mindestens fünf, sechs, sieben Jahre. Und jetzt wurde das Ganze neu aufgestellt und dann wurde das Ganze wieder eingespielt in dieses neue, also sie haben das neu gemacht. Und dann sind wir draufgekommen, dass da manches einfach nicht mehr passt. Also dass das falsch zugeordnet wurde. Das heißt, es muss man auch immer wieder kontrollieren, ob das eh ins richtige Feld gerutscht ist oder doch nicht, ja. Da muss man immer wieder aufpassen, ja. #00:35:01#

I:

Gibt es in näherer Zukunft noch Pläne in Bezug auf die Metadaten und Metadatenformate in PHAIDRA? Gibt es dann /…/ Was kommt als nächstes? #00:35:17#

B:

Also ein wichtiger, wirklich wichtiger Punkt für es ist die ORCID-ID. Die haben wir zwar momentan schon drinnen. Also man kann sie eintragen, aber sie hat noch keine Funktion. Und wir wollen sie mit einer Funktionalität hinterlegen. Das heißt, wir wollen auch, dass Forschungsdaten in der ORCID-Oberfläche hinterlegt werden können und dass sie nachgewiesen werden können. Dass die Forscher sehen, aha, sie haben so und so viele Publikationen, aber auch die Forschungsdaten, ja. Und das ist etwas, wo wir gerade dran sitzen. Das ist ein, das ist ein großer Punkt. #00:35:46#

Was ein zweiter Punkt ist, da ist ein Kollege von mir dran, so ein bisschen mit der KI zu spielen. Ja, wie kann ich Bildbeschreibungen zum Beispiel automatisieren? Hängt damit zusammen, dass wir auch immer schauen, dass wir möglichst barrierefrei unterwegs sind. Und Bildbeschreibungen sind da wichtig oder Videountertitelung oder so. Und da versuchen wir gerade mit der KI ein bisschen zu experimentieren. Ja, und außerdem, also die Arbeit an den Metadaten reißt sicher nie ab, weil irgendwie kommt dann wieder etwas Neues, ein neues Format oder irgendwie was, wo man drauf reagieren muss, ja. #00:36:21#

I:

Ja, vielen herzlichen Dank, dass Sie sich Zeit genommen haben. #00:36:26#

B:

Gerne. #00:36:26#

I:

Ich hoffe, ich habe Ihnen genug Raum gegeben, um alles / #00:36:34#.

B:

(lacht) Ich schaue gerade, ob irgendetwas nicht gesagt worden ist, aber ich glaube, ich habe jetzt alles gesagt. (…) Vielleicht noch den Punkt mit der Weitergabe, was wir weitergeben wollen. Also wir sind da sehr offen und ich habe es schon angedeutet, ich unterrichte auch in den ULGs, also in den Lehrgängen, und halte auch manchmal Vorträge und

so weiter und betreue auch Projekte bei uns, ULG-Projekte, und wir sind da sehr freigiebig. Also alles, was wir wissen, geben wir auch weiter, natürlich auch den Kolleginnen und Kollegen, die an anderen Institutionen arbeiten, also das ist kein Geheimnis. Wir sind ja auch froh, wenn wir wieder Information bekommen. Also. #00:37:23#

I:

Ja, Wissen teilen einfach. #00:37:26#

B:

Absolut. Also es gibt ja dieses RepManNet, dieses Repositorienmanager*innen-Netzwerk und da in diesen Arbeitsgruppen ist es ja auch super zu sehen, dass man eh am Gleichen arbeitet, aber halt vielleicht aus unterschiedlichen Ecken kommt und dann vielleicht sich auch Informationen holen kann irgendwo. #00:37:42#

I:

Ja und sich gegenseitig ergänzen kann. #37:44#

B: Genau, genau. #00:37:45#

I:

Ist das ein gutes Schlusswort? #00:37:51#

B:

Ja, ich glaub netzwerken ist immer ein gutes Schlusswort. #00:37:54#

I:

Dann vielen Dank. #00:37:57#

B:

Sehr gerne, wirklich sehr gern. #00:37:59#

I:

Ja, das freut mich sehr, dass Sie da waren. #00:38:03#

B:

Ja, ich freue mich auch, dass ich gefragt wurde (lacht). #00:38:07#

Anhang 2.3. Transkript des Interviews mit Rastislav Hudak

Interviewpartner: Rastislav Hudak

Datum: 30.01.2024

Ort: Wirtschaftsuniversität Wien

I:

Sehr geehrter Herr Hudak, vielen Dank, dass Sie sich heute Zeit genommen haben für das Interview. Ich freue mich sehr, dass ich Ihnen die Fragen zu meiner Masterarbeit stellen kann. Zunächst möchte ich Sie bitten, sich kurz vorzustellen. Wie heißen Sie, wo arbeiten Sie und was genau ist Ihr Aufgabenbereich? #00:00:19#

B:

Also ich heiße Rastislav Hudak. Ich arbeite an dem Zentralen Informatikdienst der Universität Wien, bin der Teamleiter von dem Team Datenmanagement unter Herr Ganguly und die Abteilung heißt IT Support for Research. Und die Aufgabe ist die Weiterentwicklung und der Betrieb von PHAIDRA, meistens. PHAIDRA Repository, das ist das Repository für Langzeitarchivierung von Digital Assets an der Universität Wien, mittlerweile auch an mehreren anderen Universitäten, Organisationen. Und wir haben dann auch noch andere Services, die wir betreiben. Im Kontext von quasi Research, IT Support for Research oder von Langzeitarchivierung, ob es für Software ist, Literaturdatenbank, Vocabulary Server und, und / #00:01:13#

I:

Also alles rund um PHAIDRA? #00:01:14#

B:

Rund um PHAIDRA, manchmal auch quasi parallel. Es gibt ein System, basierend auf EasyDB zum Beispiel, das ist dann UNIDAM, das dann halt mehr Richtung Bilder geht, wo man sich anders wieder die Metadaten selber gestalten kann, ein viel geschlosseneres

System (unv.), weil PHAIDRA ist wirklich gerichtet auf Veröffentlichung, Langzeitarchivierung. Das ist nicht immer das, was man braucht, und wir haben auch andere Lösungen, die die Forscher unterstützen. #00:01:51#

I:

Ja, ich glaube, UNIDAM wird viel von der Kultur- und Sozialanthropologie verwendet? #00:01:55#

B:

Ja, die Kunsthistoriker. Meistens, glaube ich. #00:01:58#

I:

Wie wichtig ist bei dieser Arbeit die Arbeit mit Metadaten? #00:02:00#

B:

Ist schon sehr wichtig, weil es eben der Kern von Data Management ist, dass man die Daten beschreibt und dadurch die irgendwie organisieren kann, zugänglich machen kann, auffindbar machen kann. #00:02:23#

I:

Haben Sie auch bei der Entwicklung der Metadatenformate in PHAIDRA mitgearbeitet? #00:02:35#

B:

Ja, an dem neuen Metdatenformat schon. Das haben wir mit einer Kollegin, Anna Bellotto, die ist jetzt nicht mehr in dem Team. Ich glaube, die ist jetzt bei Europeana oder so. Das haben wir uns dann zusammengesucht. Aber das kann ich dann vielleicht näher erklären später. #00:02:59#

I:

PHAIDRA ist jetzt schon um die 15 Jahre alt. Wie lange sind Sie im Team dabei? #00:03:04#

B:

Ich bin im Team seit 2011, glaube ich. PHAIDRA wurde 2000, wenn ich mich nicht täusche, 2007 gestartet und 2008 war die erste Version da. Also von dem / Ganz am Anfang war ich nicht dabei, als dieses ganze Konzept erstellt wurde und das erste Metadatenformat und so weiter. Wir sind dann gekommen und haben das weiterentwickelt, an den anderen Institutionen ausgerollt und / #00:03:35#

I:

Wie hat sich in dieser Zeit, seit Sie dabei sind, die Arbeit mit den Metadatenformaten und PHAIDRA entwickelt. Welche Veränderungen gab es da? Welche Phasen gab es da? #00:03:49#

B:

Am Anfang war glaube ich gedacht, dass man in PHAIDRA das LOM-Schema verwenden würde. Das hat man dann allerdings / Einerseits glaube ich, das LOM-Schema hat manche Sachen nicht ganz im Detail definiert, wie okay, man / in dem Lifecycle von dem Objekt definiert man Rollen und die Akteure, aber jetzt nicht genau, welche Felder man / da war nur ein E-Card oder sowas wie so ein komischer Standard als Beispiel verwendet. Also die haben / Die ursprünglichen Entwickler haben dann Metadatenfelder dazu entwickelt, auch nach Bedarf von Forschern und dadurch haben die das dann UWMETADATA genannt. #00:04:39#

Und das war schon mal das erste, ich will es nicht Problem nennen, aber eigentlich war es dann nicht mehr wirklich ein Standard, weil das unser eigener Standard war. Und dann waren die ersten Überlegungen, glaube ich, / kommen wir dann zu einem echten Standard? Und wie? Und das wurde dann weiter dadurch motiviert, dass wir im Backend / Also das Repository, das wir verwenden, heißt Fedora Commons. Ich weiß nicht, jetzt heißt es mittlerweile nur Fedora Repository oder so. Und das ist von Version 3 auf 4 auf eine andere Metadaten-Struktur, eigentlich ein komplett anderes Konzept, gegangen. #00:05:38#

Die Fedora 3, auf der wir noch immer sind, war quasi XML-basiert, da haben wir uns dann überlegt, wie wir dann dazu kommen. #00:06:17#

Wie mappen wir das? Und wir haben gesehen, dass eine Community rund um das Fedora-Repository, die sich Samvera nennt, mittlerweile, vorher war es eine Software, glaube ich, Samvera. Jetzt ist das die Community und die Software heißt Hyrax, glaube ich. Und da sind einige Universitäten dabei und / glaub um die 50 oder / weiß ich nicht mehr. Aber die haben MODS-Metadaten in Fedora. Und die haben dann ein Mapping von MODS auf RDF erstellt, um auf Fedora 4 zu kommen. #00:07:04#

Und wir sind in der Zeit gleichzeitig auch mit einem Projekt gestartet, in dem es um die Digitalisierung von Karten gegangen ist. Die Karten waren katalogisiert und aus dem Katalog konnte man relativ leicht auf MODS mappen. Und deswegen haben wir im Rahmen von diesem Projekt, und weil dieses Projekt nicht die Metadaten in PHAIDRA je editieren wollte, das musste immer upgedatet sein von dem Katalog, das ist sozusagen source of truth, haben wir uns gedacht, dann implementieren wir in PHAIDRA eine Unterstützung für MODS als ein Format, das PHAIDRA versteht und anzeigt. #00:07:51#

Man kann es nicht editieren, dadurch ist es ein viel kleineres Projekt, als wenn man es wirklich unterstützt als ein Eingabeformat. Und das heißt, wir waren schon mit MODS ein bisschen vertraut und gleichzeitig gab es dieses Projekt von Samvera, zum Mapping von MODS auf RDF. Und dann haben wir überhaupt gedacht, dann mappen wir doch UW-Metadaten auf MODS und kommen dadurch auf das RDF über dieses Mapping. #00:08:32#

Das haben wir / Dann hat sich unsere Kollegin Sonja Fiala, zurzeit glaub ich an der Fachbereichsbibliothek Philosophie oder sowas, dessen angenommen, das auch sehr gut gemacht. Wir haben uns das angeschaut und angefangen, zu überlegen, wie wir dann mit dem RDF, das dann rauskommt, in PHAIDRA umgehen können. Und haben aber dann gesehen, dass es doch, dass dieses Mapping von MODS auf RDF oft sehr vereinfacht hat das Ganze und / Was nachvollziehbar ist, wenn man bedenkt, wie viele Institutionen da mitgewirkt haben, hat man da einfach sehr viel gestrichen. Als Beispiel, wenn man eine

Person als einen Autor zum Beispiel definiert, kann man nicht sagen, was ist der Vorname und Nachname, es ist einfach ein String, ein /. #09:27#

Wir haben das schade gefunden, das jetzt quasi so wegzuschmeißen, die Informationen, die wir schon haben, so abgeteilt, und haben uns deswegen gedacht, okay, da und da werden wir uns anpassen. Und dann haben wir eigentlich aber angefangen, dieses RDF-Metadatenformat neu von Null zu gestalten. Und so sind wir weitergegangen, da war wieder / Sehr oft ist es im PHAIDRA-Bereich oder war es so, dass immer eine neue Entwicklung von einem Projekt angestoßen wurde, das das gebraucht hat. Weil einerseits hat man dann ein Feedback, ob man das tut, was die Leute auch brauchen. Andererseits sind die Ressourcen dann da, um an denen zu arbeiten. Also solche Synergien, das war schon dieses MODS mit den UB-Maps – so heißt ja das Projekt, wo man die Karten digitalisiert hat. Das Projekt, wo wir ins RDF tiefer gegangen sind, das ist die ethnographische Datenarchivierung von (unv.) #00:10:38#

I:

Kultur- und Sozialanthropologie? #00:10:40#

B:

Genau. Dieses Institut. Und die haben / Weil die sind gekommen und hatten Ideen zu sehr detaillierten Metadaten. Wobei Metadaten, die sind ziemlich komplex, weil man unterscheiden will, ob man über das digitale Objekt jetzt spricht oder das Digitalisierte oder was abgebildet ist auf dem Digitalisierten. Und das war teilweise möglich in UW-Metadaten. Aber da bräuchte man so viele neue Felder und weiß nicht was, dass man sich gedacht hat, da machen wir es mit dem neuen Metadatenformat, entwickeln wir es so schnell wie möglich weiter. #00:11:21#

Und wir haben dann einfach von (…) / Also dieses Mapping von MODS auf RDF haben wir nur so als Beispiel, als Orientierungsbeispiel genommen und das Mapping von UWMETADATA auf MODS, um zu sehen, welche Felder wir gerade momentan haben, welche die Leute verwenden. Und haben dann das Objekt aber, das Metadatenformat, oder das Schema einfach neu entwickelt. #00:11:54#

Und ich glaub, dass dann / Und im Endeffekt was, wie das jetzt dann funktioniert ist: Man hat dann die Möglichkeit gehabt, in einer neuen Version der Oberfläche dieses neue Metadatenformat zu verwenden. Und (…) aber ein Mapping von UWMETADATA auf dieses RDF wäre nach wie vor relativ schwierig und aufwendig, auch mit dem Umweg über MODS. Das heißt, das haben wir dann gelassen. Und mit der neuen Oberfläche unterstützen wir sehr wohl auch das alte Metadatenformat. Man kann nach wie vor die Objekte editieren, aber dadurch, dass das Submit quasi nicht mehr mit den neuen, äh, mit den alten – beim Submit erstellt man nicht das alte Metadatenformat, sondern das neue – werden nicht mehr wirklich neue Objekte mit dem alten Metadatenformat reinkommen. #00:13:00#

Aber eine / Es gibt momentan keinen Bedarf, alles auf das neue Metadatenformat umzustellen. Weil es aufwendig ist. Und es bringt relativ wenig, weil den Benutzern ist es relativ egal, die sehen die Metadaten sowieso. Und die anderen Repositorien kennen sich wenig genug mit UW-Metadaten wie mit dem RDF aus, weil meistens will man ein Dublin Core sehen, oder ein bestimmtes Metadaten-Schema wie OAI-PMH, OpenAIRE oder DC Terms und auf die müssen wir immer exportieren. Das heißt für die Welt da draußen ist es auch egal, welche Metadatenformate verwendet werden. Oder anders gesagt, ist es egal, dass ein Teil des Repositories dieses Format hat und ein Teil dieses – intern, weil ich will sowieso ein anderes bekommen. #00:13:53#

I:

Wie machen Sie das, wenn dann jemand kommt und Metadaten möchte, exportieren Sie dann einfach die Metadaten und schicken Sie sie zu oder was passiert dann? #00:14:01#

B:

Es kommt darauf an, welche Schnittstelle er verwenden will. Es gibt nur ein, glaube ich und auch das nur bei uns, eine Applikation oder externes Repositorium, das was anderes als OAI-PMH verwendet. Und das ist OER-Hub, da geht direkt / Das sind, das ist also der Hub für offene Bildungsressourcen. Der geht direkt in die Suchmaschine von PHAIDRA, aber normalerweise geht man über den OAI-PMH-Provider und da ist Dublin Core drinnen und oai_openaire. Und das oai_openaire nimmt sich den OpenAIRE. Ich

glaube die anderen verwenden das nicht, dass sie so ein DataCite-basiertes / ja, eigenes Flavour davon. Und Dublin Core versteht dann fast jeder. Das heißt, meistens holen sich die Leute so die Metadaten. #00:15:01#

Ich glaube / Es gibt dann noch quasi docked application, haben wir das eine Zeit genannt. Das sind so alternative Frontends für PHAIDRA oder Applikationen, wo man ein bisschen anders etwas in PHAIDRA hochlädt, ob es das institutionelle Repositorium ist, ob es u:theses ist, wo die Hochschulschriften hochgeladen werden. Die arbeiten dann natürlich direkt mit dem neuen Metadatenformat oder mit dem alten, aber die entwickeln mehr oder weniger wir in den Computerzentren. Also das ist jetzt nicht so eine Geschichte, wo man jemandem die Metadaten übergibt. #00:15:44#

I:

Okay. (…) Wir haben damit eigentlich Themenblock 1 im Wesentlichen (…) abgearbeitet. #00:16:11#

B:

Ja, ich / Leider, ich kann es so erklären, aber nicht so ganz an die Blöcke anpassen. #00:16:17#

I:

Nein, nein, nein. Es passt. Es passt wunderbar. Für mich ist es nämlich sehr spannend, weil es genau die Informationen sind, auf der / wo ich auf der Suche war. Genau. (…) Bei den Veränderungen, die Sie gerade beschrieben haben. Ich frage einfach noch kurz dazu, gab es da auch Metadatenkategorien, die neu aufgenommen werden mussten oder waren / Ich denke zum Beispiel an die Barrierefreiheit. Ich könnte mir vorstellen, dass das nicht immer / #00:17:00#

B:

Das ist noch in Arbeit mit der Barrierefreiheit, da waren wir gerade (..) im Austausch mit den Kolleginnen aus der Bibliothek, um diese Metadatenfelder hinzuzufügen, die dann

beschreiben, ob es da irgendwelche Risiken gibt oder ob es Untertitel gibt oder was auch immer. Das ist noch nicht, die sind noch nicht drinnen. #00:17:24#

Neue Metadatenkategorien / also wir haben die Metadaten jetzt mal dadurch, dass es RDF war / Das neue Metadatenformat basiert auf RDF und wurde von diesem Mapping von MODS auf RDF inspiriert und was man da eigentlich macht, das habe ich später erst gefunden, das heißt SHACL Shape oder SHACL. Das ist so ein Constraint Language Standard, kann man sagen, mit dem man definieren kann, wie der eigene RDF-Graph aussieht, welche Prädikate man denn verwenden kann. #00:18:12#

Weil (…) es ist oft so, man macht RDF, heißt, man macht eine Ontologie und macht sein kleines Feld, rund um seine Thematik. Das funktioniert dann prinzipiell sehr ähnlich wie die XML-Metadaten-Standards wie MODS, oder UWMETADATA oder LOM. Wo man dann doch immer wieder herausfindet, ein Metadatenfelder fehlt mir noch, muss ich jetzt eine neue Version von dem Ganzen machen. Man muss es mappen, weil mein Titel ist in dem Feld, dein Titel ist in einem anderen. Hier macht man dann das so, man nimmt sich die Metadatenfelder, aus den verschiedenen Standards. Und zwar ist es praktisch und flexibel, weil man (…) kann die einfach nehmen. Aber auch die Idee dahinter ist diese Linked Data-Idee, dass ich nicht wieder einen neuen Namen für den Titel hab, sondern ich nehme den, den die meisten Leute kennen. Das heißt, ich sag „Dublin Core Title“ und das sollte dann schon sehr vielen anderen Maschinen etwas sagen. #00:19:20#

I:

Und man ergänzt dann einfach, was man braucht, aus unterschiedlichen Standards, quasi? #00:19:29#

B:

Genau. Also wir haben momentan, glaube ich, mehr als 50 sogar Ontologien drinnen. Schema, BIBFRAME, ich weiß nicht, EBUCore, für diese Dinge wie MIME type oder so, und (…) RDA auch, für exotischere Sachen, die haben sehr viele Prädikate. Und die / Weil, Sie haben nach Kategorien gefragt, also jedes Feld haben wir von irgendwoher genommen, so wie wir es gebraucht haben. Am Anfang natürlich die Basis, dann halt

gleich konzipiert: Wir brauchen die bibliografischen Metadaten und Titel, wie machen wir die Personen, Identifikationsnummer, und / #00:20:12#

Die / eine Kategorie, die ist momentan, glaube ich, ein bisschen schon erweitert im Vergleich mit den UWMETADATA sind die Metadaten über das repräsentierte Objekt, weil sie eben in diesem Projekt mit den Ethnographen waren und die haben wirklich manchmal physische Sachen fotografiert, die man dann beschreiben will: Wie groß das ist. Und andererseits gab es damals auch das Projekt mit der Mediathek der Romanistik unterwegs und die sind mit digitalisierten DVDs gekommen und das war zum Beispiel eine neue Kategorie an Metadaten, weil Metadaten wie Target Audience, die sagen: Ist das ab 16 oder so? Das haben wir vorher nicht gehabt, das sind komplett neue Felder. Was gibt es noch? Inhaltsverzeichnis, Supplement, Material oder wie heißt es, dass es Extras gibt auf der DVD oder Bonusmaterial und so, dass man das alles quasi reinschreiben kann. #00:21:28#

I:

Also, das bedeutet eigentlich, egal was kommt, Sie können aus bestehenden Ontologien oder Schemata Sachen für sich holen und auf die Bedürfnisse der Nutzer*innen reagieren und die Metadatenkategorien anbinden. #00:21:44#

B:

Das können wir. Es ist immer quasi ein bisschen Entwicklungsarbeit dahinter, falls das Feld komplexer ist. Wenn es nur ein Feld ist, wie zum Beispiel diese Kategorie von Metadaten für Barrierefreiheit. Die sind relativ leicht, weil man sagt, ich habe so ein Metadatenfeld, das könnte diese 5 Werte haben. Ein Metadatenfeld, das eine Auswahl ist, ist natürlich schon drinnen und ich setze nur das Vokabular rein und das war es. Das ist, das ist nicht ein Problem. Wenn man mit sowas kommt wie, ich will ein Projekt beschreiben, dann hat man eine Metadatenfeld-Gruppe „Projekt“, wo dann drinnen steht, ja, man braucht Name, Akronym, Homepage, Identifikator, Förderer. Das ist schon komplexer. #00:22:34#

I:

Werden die Metadaten auch versucht zugänglicher zu machen, um mehr Nutzer*innen zu erreichen? #00:22:52#

B:

Benutzer*innen, also die Benutzer*innen, die PHAIDRA quasi / #00:23:02#

I:

Die Interesse zum Beispiel an den Forschungsdaten hätten oder an den Forschungs-Metadaten. #00:23:09#

B:

Ich glaube, was da vielleicht hilft, ist, wenn man kontrolliertes Vokabular verwendet, wenn man DOIs draufsetzt und, dass es über die OAI-PMH-Maschine dann in den anderen Repositorien landet, wie BASE oder EBSCO, OpenAIRE, Europeana, Google Scholar natürlich. Dass das hilft, wenn ich Forschungsdaten finden will. Meistens dann / Das Wichtigste sind die Dinge, die man von Anfang an da / wie ein vernünftiger Titel, eine Beschreibung und diese Basis-Dinge. #00:24:11#

Dass wir da jetzt in dem neuen Metadatenformat etwas Besonders / Ob das irgendwie hilft, das besser an den Benutzer zu bringen, das würde ich fast nicht sagen. Es ist etwas leichter persistente Identifikatoren für alles Mögliche zu verwenden. Also nicht nur ORCID für den Autor, sondern auch die Institution, die Affiliation, oder so, könnte einen persistenten Identifier haben davor, oder das Projekt, alles Mögliche. Im Endeffekt ist die Idee von Linked Data auch, dass RDF dann / Die Benutzerin, die uns mehr und mehr interessiert, ist die Maschine. Desto besser kann es eben in die anderen Systeme reinfließen, wenn die Maschinen das (…) das maschinenlesbar ist. Und da ist es vorbereitet, würde ich mal sagen. Dieser große Boom ist irgendwie noch nicht gekommen. #00:25:31#

Also man hat schon Success Stories natürlich von den Linked Data. Aber ich glaube, wir organisieren mit dem Metadaten-Standard jetzt die Metadaten so, dass ich nicht sehe, dass wir das irgendwie anders, besser machen könnten. Schauen wir, ob es dann in Zukunft wirklich die Ergebnisse bringt, die wir uns erhoffen, ja. #00:26:00#

Aber, wie gesagt, an die Forschungsdaten / wichtig sind da sehr oft nicht-technische Aspekte: Ist das Repository bekannt? Wissen die Leute, sie sollen dort danach suchen? Und so. Da helfen die Metadaten jetzt nicht so wahnsinnig viel. Es ist auch / Aber es ist ein guter Punkt, weil eigentlich sind die Metadaten dafür da, dass man es auffinden kann und irgendwie ein bisschen in den Kontext setzen kann. #00:26:31#

Bei diesem Projekt, das die Ethnographen – oder manchen anderen – sieht man, man versucht fast, das Objekt fast wissenschaftlich zu beschreiben und da müssen wir manchmal ein bisschen bremsen und sagen: Da bildet man jetzt nicht die Realität ab, sondern spricht über das digitale Objekt. Rollen, zum Beispiel, ist oft das Thema. Ich brauche die Rolle „Tänzerin“ oder so. Ja, kann man hinzufügen, aber die Rolle „Schriftstellerin“ brauche ich nicht, weil die Autorin von dem Objekt ist eine Autorin, und was die für einen Beruf hat; ich beschreibe nicht die Person, ja, aber manchmal ist es natürlich schon / Ich verstehe, woher es kommt. Ja, wenn ich begeistert bin, von den Objekten, will ich auch sagen, dass der Autor da und dann geboren ist und / Aber da müssen wir schon aufpassen und sagen immer, die Metadaten sind da, wegen der Auffindbarkeit und wenn man sehr detailliert über ein Objekt sprechen kann, muss man dann auch aus Urheberrechtsschutz-Gründen das dann in ein PDF verpacken. Das darf man wieder hochladen als digitales Objekt und das ist dann nicht wirklich die Aufgabe von den Metadaten in unserem Repositorium. #00:27:50#

I:

Kommen wir zum Themenblock Interoperabilität. Ich glaube, Sie würden mir nicht widersprechen, wenn ich sage, Interoperabilität ist wichtig, mit anderen Metadatenformaten. Und ich würde Sie jetzt einfach gerne fragen, wie wird das am besten gemacht? #00:28:18#

B:

Ich glaube, am besten sind da die Linked Data und die Idee von quasi einem globalen Datengraph, wo man dann wirklich auf diese grammatische / Also das RDF basiert auf diesem total primitiven Kern von Grammatik, Subject Predicate Object, und mit dem baue ich den Graph und zeige dann, dass er was mit dem Begriff von Wikipedia da hat oder

was alles dazugehört. Und die Computer können dann die Beziehungen auswerten, und das ist für die Interoperabilität / Also kommt drauf an, wie man es genau meint. Aber ich glaube, das ist das. Da geht es genau darum, dass man die Metadaten so gestaltet, dass es einerseits / Jedes RDF funktioniert auf dieselbe Weise, auf diese Triples. Und ich verwende Links und Identifikatoren, und die sind global und unique und persistent. Das heißt, zeigen nicht morgen auf etwas Anderes und dadurch mache ich die Metadaten sozusagen automatisch verständlich für alle anderen, auf diese Weise. Das kann man mit normalen Standards aber nicht wirklich erreichen. #00:29:52#

Dublin Core war sehr erfolgreich, aber auch nur deswegen, weil es extrem vereinfacht ist. Und das ist nicht ein Format, in dem man wirklich die Metadaten speichern kann, weil da kann man nicht / da kann man nur sagen „creator“ oder „contributor“ und da sagt der Bibliothekar, das geht nicht so, da brauche ich schon ein bisschen mehr. Es ist auch oft dann, wenn wir mappen sehr wohl die Frage, was ist jetzt „creator“, was ist der „contributer“, wenn wir die Rollen mappen, ja, und da ist ein „photographer“, er ist bei einem Objekt nur ein „contributor“, bei einem anderen ist er der „creator“ von der Fotografie und das kann man dann nicht wissen. #00:30:28#

Und, es gibt diesen Spruch, die Standards sind wie Zahnbürsten, dass jeder denkt, es ist eine super Idee, einen Standard zu verwenden, aber niemand will den Standard von den anderen verwenden. Und deswegen / Ich kann mir nicht vorstellen, dass es irgendeine Entität gibt, so auch die Library of Congress oder wer auch immer, der sagt, das ist jetzt der Standard und alle werden diesen Standard verwenden. Das kommt nicht. Und deswegen ist dann immer etwas nicht interoperabel und da hätte ich keine bessere Idee als das RDF und das Linked Data. #00:31:04#

I:

Also, alle Bemühungen, Standards zu verwenden sind nicht, müssen nicht unbedingt scheitern, aber es wird am Ende immer noch so viele unterschiedliche Standards geben? #00:31:16#

B:

Ja, also ich, ich kann mir vorstellen, dass es für bestimmte Bereiche dann absolut in Ordnung ist, dass es auch eine konkrete Ontologie gibt, die beschreibt dieses, was auch immer. Aber für so generelle Metadaten wie in Repositorien, wo es dann um Interoperability geht. Ja, das wäre schon schwierig. #00:31:38#

I:

Was wurde bisher gemacht? Also jetzt ist man ja auf dem Weg Richtung Linked Open Data, und wie wurde Interoperabilität bisher ermöglicht? Was hat man da gemacht? #00:31:50#

B:

Ich glaube, da hat / Also mit Interoperabilität meint man zum Beispiel, dass ich Objekte von einem anderen Repository in meinem Repository anzeige, oder so. Ich glaube, da hat wirklich eine große Rolle Dublin Core gespielt. Außer, dass die zwei Repositorien wirklich zufällig, würde ich mal sagen, das gleiche Metadatenformat verwenden. #00:32:12#

Es gibt natürlich spezielle Systeme wie Kataloge oder Archive, die verwenden dann ihre, haben eine Welt von Standards und das Bild ist wieder ein anderes. Und da ist, ich weiß nicht, MARC ist, glaube ich, bei Katalogen sehr prävalent. Also da kann dann die Interoperability auf diesem Level über MARC funktionieren. Aber kommt man zu PHAIDRA, versteht man MARC wirklich wieder nicht und das ist wirklich vielleicht das / Dass wir als Repositorium oder auch andere Repositorien, die auf diese sehr, sehr / so konzipiert sind, dass wir uns auch nicht spezialisieren, weil das wirklich schwierig ist. Also, in so einem Bereich wie einem Archiv, da kann man sagen ISAD(G) oder EAD und das wird schon ein anderes Archiv wahrscheinlich verstehen. Aber ein anderes Repositorium wahrscheinlich nicht. Also, wenn es kein Repositorium ist, das auf Archive spezialisiert ist. #00:33:11#

I:

Und gab es die Idee, also, die in PHAIDRA verwendeten Schemata anderen Repositorien zugänglich zu machen. Und einfach zu sagen, „Hey, wir verwenden das hier. Wollt ihr das auch verwenden?“ Gab es sowas? Gab es da Kooperationen? #00:33:10#

B:

Ich glaube, nein. Ich glaube, das UW-Metadatenformat, vielleicht hat man das nicht wirklich sehr gut genannt: Metadaten der Universität Wien. Warum wollte man niemand? Das war bei uns jetzt nicht wirklich der Fall. Ich / MODS von der Library of Congress aber, das hat sich dann ziemlich gut durchgesetzt. Es gibt einige Repositorien, glaube ich, die das verwenden. Bei uns? Ich glaube nicht. Jetzt, mit dem neuen Metadatenformat kann man sagen, das ist kein unser eigenes Metadatenformat. Das ist nur unsere Auswahl von den Metadatenformaten. Dadurch ist es dann nicht wirklich mehr etwas, das wir pushen als unser Metadatenformat. #00:33:22#

I:

Ich würde Sie noch fragen. Wie / Die digitalen Objekte in PHAIDRA werden immer von den Nutzer*innen beschrieben / #00:34:31#

B:

Nicht immer, weil manchmal kommen die von einem anderen System, aber jemand hat es schon im Endeffekt verzeichnet oder irgendwie katalogisiert, oder die Metadaten erfasst, das schon, aber nicht immer der Benutzer von PHAIDRA direkt. #00:34:50#

I:

Ja, wenn sie aus dem Katalog kommen, dann hat es wahrscheinlich ein Bibliothekar gemacht. Ich könnte mir vorstellen, dass die unterschiedlich beschrieben sind, also dass die Nutzer*innen, wie Sie vorhin auch gemeint haben, die Ethnologen zum Beispiel, anders beschreiben, als es eine Bibliothekarin machen würde. Gibt es da Bestrebungen oder Ideen das anzugleichen, diese Unterschiede? Oder sagt man einfach, wir lassen das? #00:35:18#

B:

Ich überlege, ob es irgendein / Man hat, glaube ich, eine Idee gehabt, wie man das verwenden sollte, aber immer haben das die Benutzer*innen auch für sich selber interpretiert

und das wäre fast hoffnungslos, das irgendwie zu versuchen. Es ist oft der Fall, dass manche kommen und sagen, das ist eine super Oberfläche, intuitiv, ich verstehe alles, und die anderen sagen, ich weiß nicht, was man hier von mir will. Und es ist einfach das Problem, dass das schon eine relativ komplexe Problematik ist, das Ganze mit den Metadaten. Manchmal geht es Richtung, eben Forschung, manchmal geht es Richtung bibliographische Metadaten. Man muss verstehen, was ist eine Ausgabe, was ist eine Reihe, was / wo gehören jetzt die Metadaten hin. Wenn ich eine Ausgabe hochlade, eine ganze als PDF, dann schreibe ich ins Ausgabe-Feld was rein. Und wenn ich einen Artikel hochlade, der in einer Ausgabe erschienen ist, dann muss ich es über das Metadatenfeld „erschienen in" und da eine Ausgabe. Und das sind so Dinge, wo, wenn man es sich mal so näher anschaut, ist das doch relativ komplex. #00:36:36#

Und unsere Schnittstelle zu Benutzer*innen ist (…) hauptsächlich diese Submit-Form, diese Eingabemaske, in der Eingabemaske kann man jetzt nicht das alles erklären, das ist fast unmöglich. Dadurch kommt immer / manchmal sagt jemand, ich verstehe es nicht, manchmal verwendet man es einfach falsch. Ich weiß nicht, da kann man sich vielleicht überlegen, vielleicht, also wir machen schon Schulungen, ob man Beispiele irgendwo wirklich ausgearbeitete vorbereitet oder sowas. #00:37:17#

Ansonsten (…) was wir haben zum Beispiel ist, man hat eine Liste von Metadatenfeldern und bei jedem Metadatenfeld ist dann dokumentiert, was das ist, wie das verwendet werden sollte, manchmal so mit Anweisungen so „FAIRify your data", was man für FAIR-Prinzipien eigentlich, wie sollte man sie verwenden, das haben wir alles schon. #00:37:50#

I:

Aber es ist keine hundertprozentige Sicherheit. #00:37:55#

B:

Na weil, (…) ich glaube einfach, weil es da wirklich dann nicht so leicht ist, sich da auch einzulesen, wenn ich auch etwas zu lesen habe, da muss ich mir die Zeit finden, um das zu verstehen und / #00:38:14#

I:

Ich glaube, auch die Forschenden werden teilweise so im Stress sein, wenn sie die Sachen hochladen des letzten / #00:38:20#

B:

Ja, ja. Aber für so diese Schienen wie Hochladen von einer meiner Veröffentlichungen oder so, da geht es auch oft durch andere Workflows rein. So zum Beispiel, wenn man eine Veröffentlichung hochladen will, geht man über die Oberfläche von dem institutionellen Repositorium, da hat man wieder anders die Metadatenfelder vorbereitet, was man vielleicht in diesem Kontext dann besser versteht, was man da reinschreiben soll und was ist notwendig und was nicht. Das heißt, so kann man auch operieren. #00:39:00#

Zum Beispiel, die Mediatheken haben ein anderes Submit, wo es angepasst ist an das Filmmaterial, und so weiter. PHAIDRA an sich, als Repositorium für alles, hat dann halt / Das war dann eine relativ lange Diskussion, wie das Metadatenformat ausschauen soll. Im alten war es so: 150. Felder und man muss sich da durchklicken und schauen, was ich jetzt erfülle oder nicht. Die Benutzer sind jetzt sehr froh, dass die das nicht mehr machen müssen (lacht), nicht mehr durchklicken. Aber, also mit dem Durchklicken durch elf Reiter, oder was das war, und da haben wir immer alle Felder. #00:39:35#

Da haben wir am Anfang gedacht, wir werden das Submit immer anpassen an das Objekt, das man hochlädt. Dass man sagt, ich lade ein Bild hoch, es ist eine digitalisierte Fotografie, dann bekomme ich diese Metadatenfelder. Dann gab es 16 oder so Submit-Forms. Da hat man gedacht, das ist ur-komplex. Das haben wir intern nicht einmal verstanden, was wir tun. Gut, dann machen wir das anders. Machen wir zehn Felder, wie Dublin Core, das war es. Dann haben wir gedacht, das ist zu wenig, das sind keine guten Metadaten, das können wir auch nicht machen. Jetzt ist es so ein Mittelweg und man sagt, wir haben diese zehn Felder oder neun, meistens auch nur eigentlich Pflichtfelder. Und dann gibt es so Bereiche da unten, die ich ausklicken kann, wo ich sage, da hier sind jetzt ein bisschen mehr bibliografische Metadaten oder mehr Metadaten über das repräsentierende Objekt, mehr Beschlagwortung vielleicht mit der GND oder so. Kann ich aber auch verstehen, dass es für jemanden zu viel ist und dass es ihn überfordert. #00:40:36#:

I:

Also es ist jetzt eine beschränktere Anzahl von Metadatenfeldern. #00:40:43#

B:

Genau, man kann die Metadatenfelder hinzufügen. #00:40:44#

I:

Wie wichtig ist es denn, diese Beschreibung der digitalen Objekte, die Sie gerade erwähnt haben, durch die unterschiedlichen Metadatenfelder zu normieren und zu dokumentieren? Welche Rolle spielt es für die langfristige Arbeit mit den Metadaten? Und wie werden dann standardisierte Vokabulare, Thesauri und Taxonomien verwendet und dokumentiert? #00:41:20#

B:

Ich glaube, der letzte Punkt ist die Antwort auf die zwei. Man sollte möglichst kontrollierte Vokabulare verwenden, ich weiß nicht, was ist eigentlich der oberste, abstrakteste Term Thesaurus oder Vokabular oder Taxonomie, Klassifikation. Aber was davon das / Also streng betrachtet, wenn Linked Data wirklich funktioniert mal, wäre es ja egal, ob ich Wikidata verwende oder GND oder was anderes, was an sich schon gemappt aufeinander ist. Dann komme ich schnell auf das, was ich in Wikidata suche, weil es mit der GND bezeichnet (…), also beschlagwortet wird, und es ist eh schon verknüpft in der Authority Control von Wikipedia oder irgendwo in dem Graph und / Das ist dann, wenn man so ein Vokabular verwendet, das gut verlinkt ist und man kümmert sich um es, dann weiß ich, dass ich in 20 Jahren immer noch mit dieser ID etwas finde. Weil, wenn auch das Vokabular nicht mehr deprecated oder sonst was, hat es sicher jemand gemappt auf was anderes und ich komme wieder zu der Kategorie, die heute aktuell ist. #00:42:47#

Das ist sicher bei manchen Bereichen nicht so wichtig, in anderen wie Ethnologie, Ethnographie, die sind schon sehr vorsichtig mit Kategorien. Weil das wird immer irgendwie, hab ich den Eindruck, dort verstanden als eine politische Aussage. Wenn ich sage, da sind die Giljaken und diese Menschen wollen morgen anders heißen, dann schaut man dann schräg die Universität Wien an, dass wir immer noch Eskimos da drin haben, oder sowas.

Da muss man aufpassen, dass man etwas nimmt, das wirklich eine Zukunft hat, als Vokabular oder / Das wäre wahrscheinlich das Wichtige. Aber auch Stichworte sind dann wichtig. Also dann einfache Stichworte, die die Benutzer angeben, die vielleicht nicht normiert sind, aber Begriffe sind, die der Benutzer wichtig findet für die Wiederauffindbarkeit. #00:43:55#

I:

Und es gilt eigentlich das gleiche, was für Vokabulare, Thesauri und so weiter gilt. Die Metadaten-Schemata sollten auch gut dokumentiert sein. Und das passiert ja jetzt auch, wenn Sie Richtung Linked Open Data gehen, dass Sie Elemente nehmen von anderen und da ist eigentlich alles schon wieder drin. Sehe ich das richtig? #00:44:21#

B:

Noch einmal bitte? #00:44:22#

I:

Wir haben gerade über Vokabulare, Thesauri und Taxonomien gesprochen und die dokumentiert sind und dokumentiert werden. Und wenn man jetzt Metadaten-Schemata verwendet, sollten die auch gut dokumentiert sein? Aber, wenn ich Sie vorhin richtig verstanden habe, Sie nehmen ohnehin Elemente, sodass die Elemente eigentlich schon dokumentiert sind. Aber ist auch das, was dann das neue Konstrukt ist, ist es auch gut dokumentiert? #00:44:58#

B:

Das neue Konstrukt ist momentan dokumentiert auf einer Wiki-Seite. Ganz formal sollte es eigentlich dieses SHACL Shape sein, da sind wir noch dabei, das hinzukriegen, dass wir das so formal definieren können. Was schon deswegen gut wäre, dass wir das dann wirklich so formal validieren können. Das ist, dass es dann dem Shape sozusagen entspricht. #00:45:28#

Momentan ist jetzt Validation nicht ein großes Thema, weil man immer diese Libraries von PHAIDRA verwendet für den Submit oder fast immer und man erstellt dieses Metadatenformat sozusagen nicht selber, sondern sagt der Library: „Gib mir bitte ein Titelfeld, das ist der Value.“ Und was die Library dann für Prädikate verwendet, um das in JSON-LD umzuwandeln, um das muss sich der Benutzer nicht kümmern. Deswegen kann es nicht sein, dass er etwas Invalides reinschreibt, sozusagen. #00:46:02#

Aber es wäre schon gut, wenn man es wirklich auch so formal definieren könnte. Das ist aber auch relativ neu, glaube ich, dieses ganze Thema und wir / Ich habe auch gesehen, dass es noch nicht sehr viele Tools gibt, mit denen man dann arbeiten kann, ohne dass man es selbst entwickelt oder so. Aber das ist ein, das ist ein nächster Schritt, den wir angehen wollen. #00:46:32#

Und sonst? Dokumentiert, wie Sie sagen, es ist immer in dem Prädikat, ist immer da vorne die Ontologie, aus der es kommt, und unter dem Link, also wenn man das JSON-LD abruft in PHAIDRA, sind alle verwendeten Felder und das sind Ontologien, dann oben in dem Kontext-Teil von JSON-LD. Und da kann man über den Link idealerweise, wenn man es richtig gemacht hat, und das ist wirklich aufrufbar, also auf die Ontologie und dort ist es dann erklärt, was das Feld bedeutet. #00:47:14#

Oft ist es eine sehr, sehr abstrakte Definition, und das ist ein Punkt, wo (…) es ein bisschen die Leute gibt, die wollen, dass die Submit-Form oder die Metadatenfelder ein bisschen verständlicher beschrieben werden, wollen das oft ändern. Und ich sag dann oft, ihr könnt was dazuschreiben, aber die Definition sollte so bleiben wie die ist, auch wenn die komplett unverständlich ist, weil so wird man in Zukunft verstehen, weil ich in die Metadaten nur schreibe „rdau:P60193“ irgendwas und das ist einfach das Property, das so definiert ist. Und wenn ich den Benutzern dann erkläre, es ist was anderes. Das, was ich denen erkläre, wird morgen wieder verschwinden. Irgendein Hilfstext und es bleibt nur das, wie es damals definiert war, also offiziell, und an das müssen wir uns halten. #00:48:13#

I:

Ja, eigentlich ist fast eine doppelte Dokumentation in gewisser Art und Weise notwendig, einerseits für die Bibliothek und den ZID im Hintergrund, damit das formal einfach stimmt und dann eine / Es gibt ja bei den Metadaten-Kategorien, gibt es ja so ein Mouse-over, wo die Kategorien ein bisschen erklärt werden, und dann bräuchte man für die Nutzer*innen eigentlich noch mal eine einfachere Erklärung. #00:48:44#

B:

Ein bisschen mit Beispiel oder so, wo man sich dann wirklich sicher sein soll, dass man das reinschreibt, was da hingehört, weil das ist wirklich oft so / Entweder sind dort bibliographische Begriffe drinnen, wie /, die eine gewisse Bedeutung haben für den Experten, so wie „Ressource“. oder „work“ oder so, was im Katalog etwas heißt, aber / Und dann (…) ist auch einfach / Zum Beispiel das Wort „Objekt“, ist in dem Bereich, ich kann es in einem Satz dreimal in drei verschiedenen Bedeutungen sagen, weil ich sage sowas wie „Das digitale Objekt ist das Objekt in diesem RDF-Satz und das ist ein JSON-Objekt dann in der Implementierung“ und, aber das ist immer anders gemeint, das Objekt. Und von dem her, wenn ich dann erkläre, ein Property mit diesen Wörtern, kann man manchmal auch als Experte nur raten, was das eigentlich genau bedeutet. Aber es ist ganz genau, sozusagen formuliert, aber trotzdem nicht immer verständlich. #00:50:02#

I:

Wie werden denn die langfristige Pflege und der Erhalt der digitalen Objekte durch die Metadaten sichergestellt? Also / #00:50:16#

B:

Also, da gibt's einerseits, wenn ich das / Die große Sorge ist, dass ich am Ende irgendein Data Lake habe oder Datenfriedhof am besten. Und um das zu vermeiden, brauche ich zu wissen, welche Dateien was sind, und dafür sind die Metadaten da. Einerseits auf dieser Ebene vielleicht. Andererseits habe ich zuletzt gehört, auch von Bibliothekaren, manchmal wenn jemand etwas hochlädt, muss ich wissen, was das ist und wie ich das verwenden kann. Oder wie die anderen das verwenden können. Und da sind die Metadaten dann sehr wichtig, quasi, dass sie es dann auch jemandem erklären, der eigentlich nicht die Target Audience ist. Weil ein Wissenschaftler, der Genofond erforscht oder so, schreibt dann in

den Titel irgendwas total Unverständliches. Und wenn er nicht dann weitere Metadaten hinzufügt, die den Bibliothekaren erklären, was das ist, dann ist es schwieriger mit Zugriffseinschränkungen oder halt diesen Dingen. #00:51:38#

I:

Ja, und da gibt es einfach unterschiedlichste Arten von Metadaten, von administrativen über technische, strukturelle. #00:51:46#

B:

Ja, strukturelle sind interessant, wenn man eine Struktur hat, zum Beispiel auch in einem System, das man aufbaut, und das Repository ist quasi nur das Archiv dahinter, dass man, wenn das System an sich nicht mehr existiert, weil das nicht langzeit-gedacht wurde, aber das Archiv schon, dass man diese Struktur in dem Archiv wiederfindet. Beziehungen sind auch wichtig, um halt die Objekte zu verknüpfen, Publikationen mit Research Data oder verschiedene Versionen von Publikationen zueinander. Dass man das alles wiedererkennt. #00:52:28#

Und technische Metadaten. Also, wenn es um Sachen geht wie MIME type und so, die helfen sehr, aber oft sind dann wieder / Also, es gibt Sachen, die sind in dem Objekt selbst drinnen und dann / Also ich finde manchmal übertreibt man ein bisschen mit den technischen Metadaten. Ich brauche in diesen Metadaten nicht die Information darüber, wie groß die Datei ist, weil die Datei ist irgendwie groß und das kann ich immer herausfinden, ja? Was in der Datei drinnen ist, quasi welches MIME type das hat, das kann ich nicht so leicht herausfinden, das ist dann zum Beispiel schon hilfreich. #00:53:20#

Und administrative Daten, die sind oft nicht in diesen bibliografischen Metadaten angezeigt, aber sind drin im System, da wo dann steht, wann das hochgeladen wurde, wann das editiert wurde. Und (…) das ist auch natürlich wichtig. #00:53:43#

Und was man für die Aufbewahrung braucht, sind dann eben diese Metadaten wie „Lizenz“, „Rechteinhaber“, wer das hochgeladen hat. Ich denke dann auch, besonders (…) wichtig kann es sein, bei Objekten, die zum Beispiel eine Sperre haben, weil da kann es dann passieren, dass ich viele Jahre später herausfinde, dieses Objekt ist wirklich gesperrt,

die Person ist nicht mehr erreichbar, aber irgendwie vielleicht durch die Lizenz oder so, kann ich mit diesem Objekt noch was anfangen. Und da helfen diese Metadaten. Weil habe ich die nicht, hat man zum Beispiel keine Lizenz angegeben oder so, dann ist das verloren. Ich kann mit den Objekten, darf ich nichts mehr machen und das ist nur schade, ja. #00:54:41#

I:

Ja, das sind dann schon so rechtliche Aspekte, die / Und wie werden die Lizenzen momentan in PHAIDRA vergeben? #00:54:50#

B:

Das muss der, der das hochlädt vergeben. Aber bei jedem Objekt, das Daten hat, also Objekte wie Collections zum Beispiel haben keine Daten an sich, aber wenn ich eine Datei hochlade, muss ich eine Lizenz angeben, oder ich gebe keine an. Also ich wähle „All rights reserved“, sozusagen. In dem Fall ist es schon empfehlenswert, dass man da reinschreibt, wer die Rechte dann besitzt. Das ist überraschendweise auch in anderen Repositorien kein Pflichtfeld oft, und das ist einfach in den Objekten irgendwie implizit drinnen. Das ist aber schwierig, weil auch der, der das hochlädt, oft nicht wirklich die Rechte hat. #00:55:31#

Aber wenn man eine Lizenz auswählt, dann muss, sollte es schon / Also die Person, die die Rechte hat, das zu verantworten über diese Lizenz, weil er definiert die Lizenzen in PHAIDRA. Man kann da die offenen Lizenzen definieren, CC-BY von 4.0, 3.0., 2.0 und Public Domain Mark. Wobei ich / Da ist also ein bisschen Diskussion, ob man es überhaupt kann in Österreich, weil man kann nicht auf das Urheberrecht verzichten. Und dann / Oder Persönlichkeitsrechte, oder wie heißt es da, da bin ich kein Experte. Aber dann gibt es noch, ich glaube, wir haben die momentan nicht in der Submit-Form, aber es gibt noch Lizenzen für Software für GNU/GPL oder so, die sind eigentlich in einem Vokabular, aber für Software empfehlen wir halt in Gitlab, nicht PHAIDRA als Repositorium und das ist dann / #00:56:27#

I:

In diesem Zusammenhang gibt es auch / Wie soll ich das sagen? Einerseits die Verwaltung der Zugriffe, aber gleichzeitig auch die Überwachung der Zugriffe? #00:56:43#

B:

Das machen wir nicht. Es gibt Zugriffsstatistiken, wobei die sich (…) die sind immer ein bisschen Heuristik, weil wir ein offenes Repositorium sind, und die Daten kann man sich dort immer anschauen, also Metadaten. Und wenn das Objekt offen ist, dann auch die Daten. Und da kommt einfach ein Request und man kann auf verschiedene, also technisch quasi erraten, ob es jetzt überhaupt eine Person oder eine Maschine ist oder nicht. #00:57:24#

Und uns / Das andere Problem dabei ist, dass man einen Teil dieser Information, wenn es eine Person identifizieren kann, gar nicht langfristig irgendwie aufbewahren darf, ohne dass man es irgendwie begründet. Das heißt, wir zählen die Zugriffe. Aber ich würde niemandem empfehlen, die irgendwie zu interpretieren. #00:57:57#

Ja, weil dann / Es ist oft so, dass man irgendwelche Zugriffszahlen sehen will. Ja, und ich erkläre dann ja, aber wir / Wenn eine IP-Adresse jeden Tag kommt, ist es klarerweise / wollen Sie das trotzdem zählen? Und dann: Ist es wichtig? Ist es nicht wichtig? Früher würde man sagen, wir wollen wissen, wie viele Menschen sich das angeschaut haben. Heute würde ich sagen, es ist wichtiger, dass die Maschine kommt, und das verbreitet, die Menschen ist da nicht so wichtig einzeln zu zählen. #00:58:31#

Wenn man das alles erklärt und versteht, versteht man nur, okay dann, ich weiß eigentlich nicht, wie ich diese Zahl interpretieren kann. Sicher sollte ich es nicht so interpretieren: Ja, dieses Objekt ist populärer als dieses da, weil es mehr Zugriffszahlen hat. Nicht davon zu sprechen, dass ich selbst reinklicken kann oder dass man einfach das Notebook jedes Mal öffnet und dieselbe Seite lädt sich und ich habe es dann nicht die 20 mal angeschaut, aber ich habe den Browser nicht ausgeschaltet. Aber die haben wir. #00:59:13#

Sowas wie: Eine konkrete Person hat sich jetzt dieses PDF heruntergeladen, haben wir nicht und ich glaube, das wäre auch nicht erwünscht. Sogar die Zugriffszahlen waren am Anfang nicht erwünscht, glaube ich, sogar vom Rektorat oder so, dass man es irgendwie vergleicht oder so. Aber später wollte man dann doch irgendwelche Zahlen sehen. Und

die haben wir, aber wie gesagt, die würde ich nicht weitergeben. Also im Archiv-Bereich zum Beispielweiß ich, da ist es schon üblich, dass wenn jemand die Daten will, dann erstellt man so ein DIP-package oder so und da ist dann wirklich immer ganz genau (…) irgendwo vermerkt, die Person hat diese Daten jetzt bekommen. Das funktioniert bei uns nicht so. #01:00:06#

I:

Das bedeutet aber auch, wenn Maschinen letztendlich auf die Daten zugreifen können, sind auch die Daten maschinenlesbar. #01:00:17#

B:

Die Daten / #01:00:18#

I:

Beziehungsweise die Metadaten. #01:00:20#

B:

Die Metadaten, wenn die Maschinen also über den OAI-PMH kommen, dann lesen sie den Dublin Core. Wenn sie über die Weboberfläche kommen, da sind schon ein bisschen diese Structured Data for Search Engine Optimization drinnen, für den Google Scholar, dass er versteht, mit welchem PDF das zusammenhängt. Und da holt man sich die Metadaten von dort. Die Daten an sich kann sich die Maschine auch abholen. Ich glaube, das machen nicht viele, aber zum Beispiel Google Scholar macht das schon, glaube ich, dass die wirklich in das PDF reingehen und auch das indexieren, und so weiter. #01:01:06#

I:

Spannend. Sie haben es vorhin schon ein bisschen erzählt, was dazu geführt hat, dass verschiedene Mappings notwendig wurden nach MODS und dann nach RDF. Beziehungsweise auch die Umwege, die dann gegangen wurden. (…) Gibt es / Was sind die zukünftigen Pläne in Bezug auf die Metadaten und die Metadatenformate? #01:01:40#

B:

Also der nächste Schritt, oder was ich demnächst mit den Kollegen machen will, ist eben, dass wir formalisieren, diese Beschreibung von dem Format. Ansonsten arbeiten wir eher an der Infrastruktur als an dem Metadatenformat an sich und daran, dass man die Metadaten, beziehungsweise die Eingabemasken sich selbst gestalten kann, per quasi Instanz. Man kann es sich schon jetzt selber gestalten mit Templates. Aber oft ist es jetzt der Fall, dass ein Institut schon sagt, okay, wir brauchen eine generelle Eingabemaske, aber wir brauchen auch eine Eingabemaske für offene Bildungsressourcen, für Forschungsdaten, das sollte was anderes sein. Und dass das irgendwie so geht, dass wir nicht in jeder Instanz alles extra implementieren müssen, sondern dass man sich selbst einstellen kann, das sind jetzt die die nächsten Schritte, also die sind nicht, die haben jetzt nicht unmittelbar mit dem Metadatenformat was zu tun. Und ab und zu kommt schon ein neues Metadatenfeld, das / Demnächst sollen eben die Metadatenfelder Barrierefreiheit kommen. #01:03:20#

I:

Gibt es irgendwas, was ich vergessen habe, was Sie noch sagen wollen? #01:03:27#

B:

Das wird mir ja einfallen, wenn ich das / #01:03:30#

I:

Auf dem Heimweg. (beide lachen) Mir geht es genauso. Deswegen herzlichen Dank, dass Sie sich die Zeit genommen haben und hergekommen sind und sich bereit erklärt haben für das Interview. Es war wirklich sehr interessant, vielen Dank. #01:03:45#

B:

Gerne #01:03:46#

Anhang 3: E-Mail-Interviews

Anhang 3.1. E-Mail-Interview mit Susanne Blumesberger

Fragen versendet am 05.03.2024
Antworten von Susanne Blumesberger unter den Fragen eingearbeitet, so erhalten am 13.03.2024

I:

Liebe Frau Blumesberger,

wir haben in unserem Interview über mehrere Punkte gesprochen, an die ich gerne noch anknüpfen würde – und teilweise habe ich noch zusätzliche Fragen. Ich wäre Ihnen sehr dankbar, wenn Sie noch Zeit hätten, folgende Fragen per E-Mail zu beantworten.

Wir haben über die Metadaten-Schemata gesprochen, die in PHAIDRA verwendet werden, etwa LOM oder auch Dublin Core. Dazu stellt sich mir noch die Frage, ob in diesem Rahmen auch Application Profiles entwickelt und verwendet wurden.

B:

Ja, das wurde/wird gemacht, unter anderem für bestimmte Nutzer*innengruppen aus dem Bereich der Kultur- und Sozialanthropologie.

I:

Außerdem würde mich noch interessieren, wie bisher (also vor der Idee Richtung Linked Data zu gehen) versucht wurde Interoperabilität herzustellen? Welche Ansätze wurden da verwendet? Wurden beispielsweise Metasuchen ermöglicht?

B:

Interoperabilität wurde mittels Schnittstellen erreicht, unter anderem über die API-Schnittstelle, von der sich Harvester Metadaten holen können, wie etwa Europeana.

I:

Wurde auch versucht, das UWMETADATA-Schema in einem Repository zugänglich zu machen?

B:

Das UWMETADATA-Schema ist im Repositorium PHAIDRA zugänglich.

I:

Wurde ein strategisches Mapping von UWMETADATA in ein anderes Metadatenformat gemacht? Oder war man mit der Nutzung von Dublin Core, LOM und DataCite eigentlich schon zufrieden?

B:

Mappings werden nach Bedarf gemacht. Es kommen auch immer wieder neue Metadatenschemata dazu.

I:

Auch die Vergabe der Rechte in PHAIDRA haben wir angesprochen. Dazu würde mich noch interessieren, ob die Informationen zu den Lizenzen maschinell austauschbar sind und damit auch über Maschinen weitergegeben werden können, und wie dieser Prozess funktioniert.

B:

Ja, die Angaben über die Rechte werden an Harvester weitergegeben, beispielsweise über die API-Schnittstelle.

I:

Außerdem wüsste ich noch gerne, ob es rechtlich bedingte Zugangsbeschränkungen in PHAIDRA gibt und wie das technisch gelöst wird.

B:

Einschränkungen der Sichtbarkeit in PHAIDRA gibt es, dahinter liegt die Struktur der Universität, man kann auf die ganze Uni, auf Institute, auf Gruppen von Personen, die selbst definiert werden können und auch auf Einzelpersonen einschränken.

I:

Wir hatten auch über unterschiedliche Arten von Metadaten gesprochen, administrative, technische, strukturelle, die alle in PHAIDRA eingesetzt werden, wenn ich es richtig verstanden habe? Sehr interessant wäre es, Beispiele dazu zu bekommen und zu erfahren, in welchem Umfang das passiert. Zusätzlich würde es mich sehr interessieren, ob auch Erhaltungs-Metadaten und Recordkeeping-Metadaten eingesetzt werden.

B:

Die Metadaten sind frei einsehbar (rechts unten) Beispiel: https://phaidra.univie.ac.at/o:2045778

Anhang 3.2. E-Mail-Interview mit Rastislav Hudak

Fragen versendet am 05.03.2024
Antworten von Rastislav Hudak unter den Fragen eingearbeitet, so erhalten am 15.03.2024

I:

Lieber Herr Hudak,

wir haben in unserem Interview über mehrere Punkte gesprochen, an die ich gerne noch anknüpfen würde – und teilweise habe ich noch zusätzliche Fragen. Ich wäre Ihnen sehr dankbar, wenn Sie noch Zeit hätten, folgende Fragen per E-Mail zu beantworten.

Wir haben über die Metadaten-Schemata gesprochen haben, die in PHAIDRA verwendet werden, etwa LOM oder auch Dublin Core. Dazu stellt sich mir noch die Frage, ob in diesem Rahmen auch Application Profiles entwickelt und verwendet wurden.

B:

Was die internen Metadatenformate betrifft, das LOM-Metadaten-Schema wurde angepasst/erweitert und das JSON-LD-Schema wird oft nach Art des Bestands anders verwendet (eine andere Untermenge von Feldern, bestimmte kontrollierte Vokabulare etc.). Man kann das auch Application Profiles nennen.

I:

Außerdem würde mich noch interessieren, wie bisher (also vor der Idee Richtung Linked Data zu gehen) versucht wurde Interoperabilität herzustellen? Welche Ansätze wurden da verwendet? Wurden beispielsweise Metasuchen ermöglicht?

B:

Fremde Systeme wie Europeana oder OpenAIRE verbinden sich über die OAI-Schnittstelle, wo die oai_dc und oai_openaire-Metadaten-Schemata (bzw. Application Profiles) eingesetzt werden. So genannte "docked-applications", die sehr eng mit PHAIDRA integriert sind (utheses.univie.ac.at, uscholar.univie.ac.at, Mediatheken, etc.), verwenden dagegen direkt die internen Metadaten-Schema von PHAIDRA. 1-2 Systeme verwenden MODS, das PHAIDRA als read-only Metadaten-Schema unterstützt, was in diesen Fällen ausreichend war, weil die fremden Systeme die datenführenden Systeme sind.

Die Metasuchmaschinen, wie die beiden oben genannten Harvester (wenn ich es richtig verstanden habe), bilden sich ihren eigenen Index, aus dem jeweiligen Metadatenformat, das aus der OAI-Schnittstelle geharvested wurde. Die einzige Ausnahme ist oerhub.at, das beim Harvesting direkt in die Suchmaschine von PHAIDRA greift (Solr) und die Metadaten aus dem PHAIDRA Suchindex dann auf einen eigenen Suchindex mappt.

I:

Wir haben auch darüber gesprochen, dass das UWMETADATA-Schema nicht in einem Repository zugänglich gemacht wurde.

B:

Das UWMETADATA-Schema ist für jedes Objekt, bei dem die Metadaten mit diesem Schema gespeichert wurden, zugänglich. Es handelt sich aber um eine Erweiterung von LOM. LOM ist ein Standard, UWMETADATA an sich nicht. Vielleicht hat man das mit der "Zugänglichkeit" so gemeint: ohne weiteres wird ein fremdes System das UWMETADATA-Schema nur teilweise verstehen können.

I:

Aber wurde vielleicht versucht, ein strategisches Mapping von UWMETADATA in ein anderes Metadatenformat zu machen? Oder war man mit der Nutzung von Dublin Core, LOM und Data Cite eigentlich schon zufrieden?

B:

Im Rahmen vom geplanten Umstieg auf RDF wurde ein Mapping von UWMETADATA auf MODS definiert, aber am Ende nicht entwickelt. Die Mappings auf DC, OpenAIRE Application Profile, DataCite (und zukünftig dcterms) haben bis jetzt sonst alle Anwendungsfälle abgedeckt.

I:

Wir haben auch darüber gesprochen, wie wichtig die Verwendung und Dokumentation von standardisierten und kontrollierten Vokabularen, Thesauren und Taxonomien bei der gegenwärtigen Entwicklung des RDF-basierten Metadatenformats ist. Wie wichtig waren der Einsatz und die Dokumentation von standardisierten und kontrollierten Vokabularen, Thesauren und Taxonomien in den bisherigen Formaten und wie wurde das bisher gemacht?

B:

In UWMETADATA wurden über das (LOM-native) "classification"-Element einige Vokabulare eingebaut, wie z.B. ÖFOS, Getty, Dewey, BIC, Basisklassifikation, EuroVoc etc., indem diese in unsere eigene interne Datenbank übergetragen wurden. Zur Wichtigkeit der jeweiligen Dokumentation des Vokabulars an sich, inhaltlich, kann ich nicht viel sagen, weil ich persönlich die Klassifikationen nicht verwende. Für uns Techniker ist eher die Dokumentation von dem Exportformat, bzw. idealerweise von der Schnittstelle, falls vorhanden, natürlich sehr wichtig, da geht es aber nicht inhaltlich um das Vokabular.

I:

Auch die Vergabe der Rechte in PHAIDRA haben wir angesprochen. Dazu würde mich noch interessieren, ob die Informationen zu den Lizenzen maschinell austauschbar sind und damit auch über Maschinen weitergegeben werden können, und wie dieser Prozess funktioniert.

B:

Die Lizenzen sind in Form von einer URI angegeben, und damit direkt maschinenlesbar. Zum Beispiel http://creativecommons.org/licenses/by-sa/4.0/. Auch in einem Fall, in dem keine Lizenz angegeben ist, handelt es sich um eine URI: http://rightsstatements.org/vocab/InC/1.0/

I:

Außerdem wüsste ich noch gerne, ob es rechtlich bedingte Zugangsbeschränkungen in PHAIDRA gibt und wie das technisch gelöst wird.

B:

In der öffentlich zugänglichen PHAIDRA-Instanz kommen Zugangsbeschränkungen zum Einsatz, wenn:

* ein Embargo eingehalten werden muss
* wenn die Benutzerinnen ihre Daten nicht ohne Rücksprache weitergeben wollen, oder es sich um sensitives Material handelt
* wenn es sich um obsolete oder fehlerhafte Inhalte handelt

Ein Material, für dessen Veröffentlichung grundsätzlich keine Rechte vorliegen, darf typischerweise in PHAIDRA nicht hochgeladen werden, andernfalls wir es wieder gelöscht.

Im einem Fall, in dem es nur um Archivierung und nicht Veröffentlichung geht, kann man eine uni-interne Instanz verwenden, die aus dem Internet nicht zugänglich ist. Hier sind die Objekte typischerweise auch noch mit weiteren Zugangseinschränkungen versehen.

Technisch ist die Sperre eine Liste von Benutzernamen, Organisationseinheiten oder Gruppen, für die die Daten zugänglich sein sollen. Bei einer Sperre geht es ausschließlich um Daten, die Metadaten sind immer sichtbar.

I:

Wir hatten auch über unterschiedliche Arten von Metadaten gesprochen, administrative, technische, strukturelle, die alle in PHAIDRA eingesetzt werden, wenn ich es richtig verstanden habe? Sehr interessant wäre es, Beispiele dazu zu bekommen und zu erfahren, in welchem Umfang das passiert. Zusätzlich würde es mich sehr interessieren, ob auch Erhaltungs-Metadaten und Recordkeeping-Metadaten eingesetzt werden.

B:

Ein Beispiel für administrative Metadaten ist z.B. "ownerid", das heißt, der Account, mit dem das Objekt hochgeladen wurde. Diese Metadaten sind einerseits technisch wichtig (z.B. nur diese Person darf das Objekt bearbeiten), anderseits rechtlich (diese Person hat den Nutzungsbedingungen zugestimmt, mit dem Upload die Rechte von Dritten nicht zu verletzen). Zu den administrativen würde ich auch den „audit log“ zählen: hier wird mitgeschrieben, welcher Account was und wann mit dem Objekt gemacht hat. Auch ein Teil der bibliographischen Metadaten ist aus administrativer Sicht wichtig: Beschreibung, Lizenz oder Rechte bestimmen nicht nur, was die Benutzerinnen mit den Objekten machen können, sondern auch, was der Bibliothekar/Repository Manager damit machen kann (z.B. ob es sinnvoll ist, die Objekte an Harvester weiterzuleiten/disseminieren).

Technische Metadaten sind z.B. „MIME type“, das bestimmt, ob es sich um Bild oder ein PDF- Dokument handelt, wie groß die Datei ist, etc.: Dementsprechend wird das Repository mit dem Objekt umgehen (z.B. bei der Anzeige).

Strukturelle Metadaten erlauben, dass die Bestände z.B. hierarchisch aufgebaut werden, oder dass manche Objekte enger verknüpft sind, oder man die Objekte versionieren kann, etc. Sie sind oft essenziell, um das Objekt in seinem Kontext zu verstehen.

Die Erhaltungsmetadaten, wie z.B. eine „checksum“, zielen dann darauf, die digitalen Objekte von unerwünschten Änderungen durch bitrot zu schützen. Bei manchen Objekten gibt es auch Erhaltungsmetadaten, die beschreiben, wie das digitale Objekt entstanden ist (z.B. wie eine Fotografie digitalisiert wurde, in welchem Zustand etc.).

Zu den Recordkeeping-Metadaten kann ich nicht viel sagen, ich glaube die werden eher in einem anderen Umfeld verwendet (wie z.B. in Archiven). PHAIDRA beschäftigt sich mit digitalen Objekten und ich glaube, unter „record“ versteht man typischerweise etwas anderes, obwohl die Metadatenfelder, die man für deskriptive und obengenannte Metadatentypen braucht, wahrscheinlich großteils dieselben sind, die man beim Recordkeeping braucht.

Anhang 4: Kodierleitfaden mit Analysekategorien und Ankerbeispielen für die Auswertung der Interviews

Abkürzungen für die Expert*inne/Autor*innen:

RG	Ganguly, Raman, Expert*innen-Interview, Wien, 20.12.2023, s. Anhang 2.1.
RH	Hudak, Rastislav, Expert*innen-Interview, Wien, 30.01.2024, s. Anhang 2.3.
SB	Blumesberger, Susanne, Expert*innen-Interview, Wien, 15.01.2024, s. Anhang 2.2.
RHM	Hudak, Rastislav, Expert*innen-Interview per Mail, Wien, 15.03.2024, s. Anhang 3.2.
SBM	Blumesberger, Susanne Expert*innen-Interview per Mail, Wien, 05.03.2024, s. Anhang 3.2.

Kategorie	Bezeichnung der Kategorie	Definition	Quelle	Kriterien	Ankerbeispiel	Autor*in	Fundstelle
1.	Verwendung von Community Standards	Dieser Aspekt richtet beinhaltet die Berücksichtigung von Standards, die in der jeweiligen Community verwendet werden. Diese sind so zu verwenden, dass sie den Ansprüchen der digitalen Objekte der Sammlung, den Nutzer*innen und der Art der Nutzung der Sammlung in der Gegenwart gerecht werden. Dabei sollen auch mögliche Nutzungsarten in der Zukunft berücksichtigt werden.	vgl. NISO 2007:61				
1.1.	Auswahl des Schemas mit Bezug auf die in der Community verwendeten Schemata	Es sollten Metadaten-Schemata verwendet werden, die in der Community üblich sind, oder zumindest ein Mapping zu diesen zu erstellen, anstatt lokal entwickelte Schemata zu verwenden. Die Berücksichtigung der in der jeweiligen Community verwendeten Standards kann auch dazu dienen, Interoperabilität zu ermöglichen. Wenn Organisationen bei der Auswahl der Metadaten-Schemata untersuchen, welche in anderen, ihnen ähnlichen Organisationen verwen-	vgl. NISO 2007:63	Diese Kategorie kann sowohl in der Vergangenheit als auch in der Gegenwart verwendete Standards und Schemata sowie die geplante Verwendung von Standards und Schemata umfassen. Diese können zu unterschiedlichsten Disziplinen gehören, da das Repositorium PHAIDRA der Universitätsbibliothek der Univer-	Na, also, PHAIDRA ist ja eigentlich eine / basiert auf einer Open Source Software Fedora. Und der ZID, der Zentrale Informatikdienst bei uns, hat das weiterentwickelt. Und wir haben von Anfang an einige Metadaten-Schemata drinnen gehabt in PHAIDRA, Dublin Core zum Beispiel, und das LOM-Schema – Learning Object Metadata. Das hatten wir deshalb drinnen, weil wir auch die Lehre sehr dabeihaben wollten. Und das hat sich jetzt als sehr günstig erwiesen, weil inzwischen gibt es, also eh schon länger, Projekte zu Open	SB	#00:09:36#

Kategorie	Bezeichnung der Kategorie	Definition	Quelle	Kriterien	Ankerbeispiel	Autor*in	Fundstelle
		det werden, führt die Verwendung des gleichen Schemas zu Interoperabilität zwischen den Sammlungen.		sität Wien allen ihren Instituten zur Verfügung stellt und so unterschiedlichste Fachbereiche ihre digitalen Objekte dort hochladen können.	Educational Resources, die genau dieses Metadatenschema benötigen, um eben diese Objekte gut beschreiben zu können.		
1.2.	Verwendung von Application Profiles	Ist ein in der Community verbreitetes Schema ausgewählt, können Application Profiles verwendet werden, um es den eigenen Bedarfen anzupassen. Application Profiles dienen dazu, aus dem jeweiligen Metadaten-Standard diejenigen Elemente auszuwählen, die für die Anwendung etwa in einer Bibliothek notwendig sind. Sie werden verwendet, um zu definieren, wie die Elemente des Schemas, die mehrere Anwendungsmöglichkeiten oder Interpretationen zulassen, im eigenen Fall, der eigenen Anwendung, verwendet werden. Dieses Application Profile sollte genau festgelegt und dokumentiert werden. Durch Application Pro-	vgl. NISO 2007:63f.	Auch diese Kategorie soll sich auf die Vergangenheit, Gegenwart und Zukunft beziehen können. Zu dieser Kategorie sollen alle Arten von Interpretationen von Schemata für die eigene Nutzung im Repositorium gehören, auch wenn sie nicht explizit als Application Profiles bezeichnet werden. Dazu gehört auch die Kombination unterschiedlicher Standards.	Was die internen Metadatenformate betrifft, das LOM-Metadaten-Schema wurde angepasst/erweitert und das JSON-LD-Schema wird oft nach Art des Bestands anders verwendet (eine andere Untermenge von Feldern, bestimmte kontrollierte Vokabulare etc.). Man kann das auch Application Profiles nennen.	RHM	Zeile 14-17

Kategorie	Bezeichnung der Kategorie	Definition	Quelle	Kriterien	Ankerbeispiel	Autor*in	Fundstelle
		files können auch unterschiedliche Metadaten-Schemata kombiniert werden. Dabei können die Vorgaben des ursprünglichen Schemas modifiziert und entsprechend den eigenen Bedürfnissen ergänzt werden. Außerdem können über Application Profiles auch Elemente zu bestehenden Metadaten-Schemata hinzugefügt werden, die an der jeweiligen Organisation zusätzlich notwendig sind, und lokal entwickelt werden können.					
2.	Unterstützung von Interoperabilität	Dieser Aspekt beinhaltet die Unterstützung von Interoperabilität	vgl. NISO 2007:76				
2.1.	Vergabe von detaillierten Metadaten als Grundlage für Interoperabilität	Gute Metadaten sollten „coherent, meaningful, and useful in global contexts" (NISO 2007: 76) sein und zwar unabhängig von der Institution, an der sie entstanden sind. Dafür müssen sie alle relevanten Informationen zu den jeweiligen digitalen Objekten enthalten. Denn wenn sich die Objekte und	vgl. NISO 2007:76		Ja, selbstverständlich ist das ein wichtiger Punkt bei uns. Also unser Repositorium wurde ja 2008 sozusagen in die Welt gesetzt. Seitdem gibt es das Repositorium. Und die Arbeit an den Metadaten, also mal zu überlegen, welche Metadaten man überhaupt benötigt in einem Repositorium, das war ein ganz ein wichtiger Punkt und hat eigentlich das	SB	#00:01:56# - #00:02:48#

Kategorie	Bezeichnung der Kategorie	Definition	Quelle	Kriterien	Ankerbeispiel	Autor*in	Fundstelle
		Metadaten in neuen Kontexten finden, in der vernetzten Welt, müssen sie genauso deutbar und lesbar sein, wie an der Ursprungsinstitution, wo noch durch den Kontext implizierte Schlüsse möglich waren.			erste Jahr richtig ausgefüllt. Also wir haben 2007 begonnen. Da startete das Projekt, so hieß es damals: Projekt Digitale Langzeitarchivierung. Und wir haben wirklich ein Jahr lang versucht herauszufinden, welche Metadaten von welchen Personen gebraucht werden. #00:01:56# Also wir sind dann wirklich in der Uni Wien herumgegangen, haben uns Pilotpartner gesucht. Aus unterschiedlichen Bereichen, zum Beispiel aus der Translationswissenschaft, aus der Bibliothek, aber auch aus der Physik, anderen Bereichen, und haben einfach geschaut: Was brauchen die Wissenschaftler? Was brauchen die Forschenden, um eben ihre Objekte gut beschreiben zu können? #00:02:17# Das war ein ganz ein wichtiger Punkt, weil natürlich, wie man sich vorstellen kann, die Uni Wien ist riesig und sehr, sehr heterogen. Das heißt, wir haben unglaublich unterschiedliche Daten und die alle unter ei-		

Kategorie	Bezeichnung der Kategorie	Definition	Quelle	Kriterien	Ankerbeispiel	Autor*in	Fundstelle
					nen Hut zu bringen und zu sagen, ich kann die jetzt einfach in einem Repositorium so beschreiben, dass man wirklich das Gefühl hat, man hat die Daten gut beschrieben und das kommt eigentlich dem zugute, was man wirklich auch möchte. Das heißt, man trifft wirklich genau den Punkt, ist gar nicht so einfach, und das war für uns eine ganz wichtige Phase natürlich. #00:02:48#		
2.2.	Zusammenarbeit zwischen Organisationen und Aggregatoren für Interoperabilität	Die Metadaten der Sammlungen von Bibliotheken, Archiven und Museen sind oft nicht über Suchmaschinen auffindbar, da sie sich im Deep Web befinden. Daher erhalten die Institutionen nicht die Aufmerksamkeit der Nutzer*innen, die sie sich wünschen würden. Ein Ansatz, der dem entgegensteuern soll, ist der Einsatz von Aggregatoren (vgl. Zeng und Qin 2022: 305). Aggregatoren sind dabei "a central organization [which] takes the role of facilitating the discoverability and use of the resources, by collecting their associated	vgl. Zeng und Qin 2022: 305; Freire u. a. 2017: 425; vgl. NISO 2007: 76	Zusammenarbeit mit Aggregatoren, diese bieten zusätzliche Dienstleistungen. Aggregatoren können auch Metadaten harvesten. Der Einsatz von Harvesting-Protokollen durch das Repositorium wird in einer eigenen Kategorie, unter 2.5. analysiert.	Also das war ein wichtiger Punkt, den wir von Anfang an mitgeplant haben, dass wir eben kein Silo sind, wo die Leute was reinstellen. Und dann ist es da drinnen und dann können die nicht mehr irgendwo das weiterverbreiten. Sondern es war von Anfang an mitgedacht, dass wir ja, also einen Austausch haben wollen mit anderen Systemen. Und da waren wir recht früh auch mit OpenAIRE zum Beispiel in Kontakt. Also ich glaub seit 2009, dieses EU-Projekt, die auch harvesten, ja, die sich auch Daten holen und Europeana. Das war auch / vor allem Europeana Libraries, das	SB	#00:21:15#

Kategorie	Bezeichnung der Kategorie	Definition	Quelle	Kriterien	Ankerbeispiel	Autor*in	Fundstelle
		metadata. The central organization has the possibility to further promote the usage of the resources by means that cannot be efficiently undertaken by each digital library in isolation" (Freire u. a. 2017: 425). Aggregatoren sollten mit den erhaltenen Metadaten arbeiten und sie, wenn notwendig, überarbeiten und verbessern. Die Verwendung entsprechender Standards und die Zusammenarbeit zwischen den Organisationen, die Daten zur Verfügung stellen, und den Aggregatoren, die zusätzliche Dienstleistungen bieten, sind wesentlich, um digitale Sammlungen über Aggregatoren zugänglich zu machen (vgl. NISO 2007: 76).			war ein großes Projekt bei uns. Das uns sehr geholfen hat, weil da sehr viel digitalisiert werden konnte und andererseits natürlich auch genau dieser Austausch angeregt werden konnte. Ja, also das, das war von Anfang an ein großer Punkt. Das haben wir immer gesagt, dass das wichtig ist, ja.		
2.3.	Interoperabilität unterschiedlicher Metadaten-Schemata durch ihre Speicherung in Reposito-	Laut den NISO-Prinzipien sollen Metadaten-Schemata idealerweise in dafür vorgesehenen Repositorien abgelegt werden und dort normierte Informationen zu Identifizierung, Definition und Verwendung der einzel-	vgl. NISO 2007:77; vgl. GO FAIR 2023c	Diese Kategorie soll die Dokumentation und Zugänglichmachung der verwendeten Metadaten-Schemata in Repositorien und auch anderswo beinhalten. Dabei soll die	Das UWMETADATA-Schema ist für jedes Objekt, bei dem die Metadaten mit diesem Schema gespeichert wurden, zugänglich. Es handelt sich aber um eine Erweiterung von LOM. LOM ist ein Standard, UWMETADATA an sich nicht. Vielleicht hat man	RHM	Zeile 45-49

Kategorie	Bezeichnung der Kategorie	Definition	Quelle	Kriterien	Ankerbeispiel	Autor*in	Fundstelle
	rien und generell öffentliche Dokumentation und Zugänglichmachung	nen Metadatenelemente bieten. Dies soll die einheitliche Anwendung des Metadaten-Schemas und seiner Elemente innerhalb der jeweiligen Organisation, aber auch in der gesamten Community ermöglichen. Außerdem können so weitere Communities die Metadaten-Schemata und ihre Elemente wiederverwenden (vgl. NISO 2007:77). Diese Kategorie soll jedoch nicht nur die Dokumentation der verwendeten Metadaten-Schemata in Repositorien umfassen, sondern jegliche Dokumentation, der Metadaten-Schemata, die der Community zugänglich sind (vgl. GO FAIR 2023c)		bisherige, gegenwärtige und zukünftige Arbeit mit Metadaten-Schemata umfasst werden.	das mit der "Zugänglichkeit" so gemeint: ohne weiteres wird ein fremdes System das UWMETADATA-Schema nur teilweise verstehen können.		

Kategorie	Bezeichnung der Kategorie	Definition	Quelle	Kriterien	Ankerbeispiel	Autor*in	Fundstelle
2.4.	Interoperabilität unterschiedlicher Metadaten-Schemata durch Mappings	Interoperabilität kann auch erreicht werden, indem die Elemente eines Schemas in ein anderes Schema gemappt werden. Diese Mappings ermöglichen es Nutzer*innen, die ein Schema verwenden, ein anderes Schema zu verstehen. Mappings können außerdem dazu dienen Daten, die anhand eines Metadaten-Schemas entwickelt wurden, in ein anderes Metadaten-Schema zu übertragen. Wird ein lokal entwickeltes Metadaten-Schema verwendet, das nicht auf einem Standard basiert, schlägt die NISO vor, ein strategisches Mapping in ein Standard-Metadaten-Schema durchzuführen, um Interoperabilitätswünschen in der Zukunft gerecht werden zu können.	vgl. NISO 2007:77	Die Mappings sollen sich sowohl auf durchgeführt als auch begonnene Mappings beziehen.	Im Rahmen vom geplanten Umstieg auf RDF wurde ein Mapping von UWMETADATA auf MODS definiert, aber am Ende nicht entwickelt. Die Mappings auf DC, OpenAIRE Application Profile, DataCite (und zukünftig dcterms) haben bis jetzt sonst alle Anwendungsfälle abgedeckt.	RHM	Zeile 56-58
2.5.	Interoperabilität durch Verwendung von Harvesting-Protokollen	Harvesting-Protokolle dienen dazu, Metadaten automatisiert „abzuernten“ oder „einzusammeln“. Ein weit verbreitetes Harvesting-Protokoll, das dafür verwendet wird, ist das Open Archives	vgl. NISO 2007:78	Diese Kategorie konzentriert sich auf die unterschiedlichen Arten von Harvesting-Protokollen. Sie nennt nicht die Aggregatoren,	Es kommt darauf an, welche Schnittstelle er verwenden will. Es gibt nur ein, glaube ich und auch das nur bei uns, eine Applikation oder externes Repositorium, das was anderes als OAI-PMH verwendet. Und	RH	#00:15:01#

Kategorie	Bezeichnung der Kategorie	Definition	Quelle	Kriterien	Ankerbeispiel	Autor*in	Fundstelle
		Initiative Protocol für Metadata Harvesting (kurz OAI/PMH). Institutionen, die es verwenden, machen damit ihre Metadaten leichter zugänglich.		die Harvesting-Protokolle einsetzen, und in Kategorie 2.2. festgehalten werden	das ist OER-Hub, da geht direkt / Das sind, das ist also der Hub für offene Bildungsressourcen. Der geht direkt in die Suchmaschine von PHAIDRA, aber normalerweise geht man über den OAI-PMH-Provider und da ist Dublin Core drinnen und oai_openaire. Und das oai_openaire nimmt sich den OpenAIRE. Ich glaube die anderen verwenden das nicht, dass sie so ein DataCite-basiertes / ja, eigenes Flavour davon. Und Dublin Core versteht dann fast jeder. Das heißt, meistens holen sich die Leute so die Metadaten		
3.	Der Einsatz von standardisierten und kontrollierten Begriffen	Unterschiedliche digitale Objekte und Objekt-Gruppen sollen mit Metadaten beschrieben werden, indem nach Möglichkeit standardisierte und kontrollierte Begriffe verwendet werden	vgl. NISO 2007:79				
3.1.	Die Verwendung von standardisierten und kontrollierten Vokabu-	Analog zu den zahlreichen veröffentlichten Metadaten-Schemata, gibt es zahlreiche veröffentlichte kontrollierte Vokabulare, Thesauren und Taxonomien. Zu den Ele-	vgl. NISO 2007:79		Also wir versuchen immer mehr zu standardisieren. Also das war am Anfang vielleicht nicht ganz so, und das ist jetzt schon ziemlich stark, also wir versuchen wirklich, möglichst standardisiertes Vokabular zu	SB	#00:13:34#

Kategorie	Bezeichnung der Kategorie	Definition	Quelle	Kriterien	Ankerbeispiel	Autor*in	Fundstelle
	laren, Thesauren und Taxonomien	menten, die sich in kontrollierten Vokabularen, Thesauren und Taxonomien finden, gehören Namen von Personen, Unternehmen, Orte, Werktitel, sowie thematische und sich auf das Genre beziehende Schlagworte.			verwenden, zum Teil, was aber nicht wirklich durchhaltbar ist, weil das teilweise nicht passt.		
3.2.	Dokumentation der Verwendung externer als auch lokal entwickelter kontrollierter Vokabulare, Thesauren, Taxonomien	Häufig werden wegen der Nutzer*innenfreundlichkeit neben lokalen, speziell für die Kollektion entwickelten Ansätzen zusätzlich Vokabulare, Thesauren, Taxonomien aus veröffentlichten Standards verwendet. Dabei sollte immer die Art der Verwendung dokumentiert werden. Außerdem sollten interne Richtlinien entwickelt werden, um Metadaten-Erzeuger dabei zu unterstützen, die gewählten Begriffe richtig zu verwenden	vgl. NISO 2007:79		Ja, die werden dokumentiert und es gibt auch die Möglichkeit, selbst Vokabularien einzuspielen.	SB	#00:16:16#
4.	Festlegung der rechtlichen Nutzungsbedingungen	Gute Metadaten bieten Informationen zu den rechtlichen Nutzungsbedingungen von digitalen Objekten.	vgl. NISO 2007:81				

Kategorie	Bezeichnung der Kategorie	Definition	Quelle	Kriterien	Ankerbeispiel	Autor*in	Fundstelle
4.1.	Beschreibung der Rechte	Zur Beschreibung der Rechte gehören unter anderem urheberrechtliche Aspekte der Objekte, Informationen darüber, unter welchen Bedingungen die Rechteinhaber*innen die Verwendung der Objekte erlauben, Angaben dazu, welche Nutzer*innenarten autorisiert sind, die Objekte zu benutzen, sowie Übereinkünfte zwischen diesen beiden Parteien dazu, wie Objekte wieder- bzw. weiterverwendet werden dürfen. Ein Beispiel für Lizenzen, die Nutzungsbedingungen von Objekten erfassen, sind die Creative Commons-Lizenzen.	vgl. NISO 2007:81		Also zunächst einmal, es gibt Nutzungsbedingungen, die akzeptiert werden müssen, bevor man etwas hochladen kann. Und dort steht drinnen, dass man eben nichts hochlädt, was irgendwelchen Rechten widerspricht. Das heißt, man sagt, man hat die Rechte auf dieses Objekt. Damit ist einmal die Person, die hochlädt, verantwortlich für das, was diese Person hochlädt. Ja, das ist einmal der Ausgangspunkt.	SB	#00:26:26#
4.2.	Management und Kommunikation von Rechten	In diesem Bereich geht es vor allem darum, Rechte an Objekten zu managen. Dafür sollten Lizenzinformationen für bestimmte Ressourcen entwickelt und maschinell austauschbar gemacht werden. Zwei Beispiele für Metadaten-Schemata sind ONIX for Publication Licenses (kurz ONIX-PL, stellt Lizenzbedingungen für Bibliotheks-Abonnements in XML	vgl. NISO 2007:81		Ja, die Angaben über die Rechte werden an Harvester weitergegeben, beispielsweise über die API-Schnittstelle.	SBM	Zeile 56-57

Kategorie	Bezeichnung der Kategorie	Definition	Quelle	Kriterien	Ankerbeispiel	Autor*in	Fundstelle
		dar) und PLUS License Data Format, das Lizenzbedingungen bei Bildern austauschbar macht.					
4.3.	Verwaltung rechtlich begründeter Zugangsbedingungen zu Objekten	Um die Interaktion und Transaktion von Rechten zwischen den Inhaber*innen der Rechte und den möglichen Nutzer*innen zu erleichtern, sollten maschinenlesbare Sprachen verwendet werden. Sie sind üblicherweise in der Lage, sich auf eine Reihe von Standards zu beziehen und sind mit einem Set von Kommunikationsprotokollen kompatibel. Dies dient dazu den digitalen Workflow im Bereich der Lizenzrechte zu managen und kann die Autorisierung von Nutzer*innen, die Durchsetzung von Vereinbarungen, die Zugangskontrolle zu den digitalen Objekten und die Überwachung der Zugriffe auf die Objekte einschließen.	vgl. NISO 2007:81		Einschränkungen der Sichtbarkeit in PHAIDRA gibt es, dahinter liegt die Struktur der Universität, man kann auf die ganze Uni, auf Institute, auf Gruppen von Personen, die selbst definiert werden können, und auch auf Einzelpersonen einschränken.	SBM	Zeile 65-67

Kategorie	Bezeichnung der Kategorie	Definition	Quelle	Kriterien	Ankerbeispiel	Autor*in	Fundstelle
5.	Metadaten für langfristigen Erhalt und Pflege der digitalen Objekte	Metadaten sollen dazu dienen, die Objekte der jeweiligen Sammlung langfristig zu pflegen und zu erhalten. Die für die Erfüllung des fünften Prinzips der NISO relevanten Metadaten-Arten sind technische Metadaten, Erhaltungsmetadaten (auch Retentionsmetadaten genannt), Recordkeeping-Metadaten und die strukturellen Metadaten.	vgl. NISO 2007:83				
5.1.	Der Einsatz von administrativen Metadaten	Administrative Metadaten enthalten Informationen, die dazu dienen sollen, digitale Objekte zu managen. Diese Kategorie von Metadaten kann etwa erfassen, wann und wie ein digitales Objekt entstanden ist, wer die Zugriffsrechte verwaltet oder die Inhalte archiviert, wie die digitalen Objekte weiterverwendet und eventuell -verarbeitet wurden, ebenso welche Art von Beschränkungen bei Zugang oder der Weiterverwendung vorliegen.	vgl. NISO 2007:83		Und administrative Daten, die sind oft nicht in diesen bibliografischen Metadaten angezeigt, aber sind drin im System, da wo dann steht, wann das hochgeladen wurde, wann das editiert wurde. Und (…) das ist auch natürlich wichtig	RH	#00:53:43#
5.2.	Der Einsatz von technischen Metadaten	Technische Metadaten sind eine spezifische Form von administrativen Metadaten und dienen dazu digitale Da-	vgl. NISO 2007:83		Und technische Metadaten. Also, wenn es um Sachen geht, wie MIME type und so, die helfen sehr, aber oft sind dann wieder / Also, es gibt Sachen,	RH	#00:53:20#

Kategorie	Bezeichnung der Kategorie	Definition	Quelle	Kriterien	Ankerbeispiel	Autor*in	Fundstelle
		teien zu beschreiben. Sie erfassen, wie der Name sagt, technische Informationen digitaler Objekte: wie etwa Format und Dateigröße. Technische Metadaten sind dabei notwendig, um eine langfristige Nutzung von digitalen Objekten sicherzustellen, oder diese wiederherzustellen, falls sie beschädigt werden sollten.			die sind in dem Objekt selbst drinnen und dann / Also ich finde manchmal übertreibt man ein bisschen mit den technischen Metadaten. Ich brauche in diesen Metadaten nicht die Information darüber, wie groß die Datei ist, weil die Datei ist irgendwie groß und das kann ich immer herausfinden, ja? Was in der Datei drinnen ist, quasi welches MIME type das hat, das kann ich nicht so leicht herausfinden, das ist dann zum Beispiel schon hilfreich.		
5.3.	Der Einsatz von Erhaltungsmetadaten	Erhaltungsmetadaten sind eine spezifische Form von administrativen Metadaten. Erhaltungssmetadaten sollen den langfristigen Erhalt digitaler Objekte ermöglichen. Dafür können Datenkategorien notwendig sein, die detaillierte technische Metadaten erfassen. Aber auch Datenkategorien, die die Beziehungen zwischen digitalen Objekten und ihrem Kontext, sowie Verantwortlichkeiten für die digitalen Objekte und	vgl. NISO 2007:83		Die Erhaltungsmetadaten, wie z.B. eine „checksum“, zielen dann darauf, die digitalen Objekte von unerwünschten Änderungen durch bitrot zu schützen. Bei manchen Objekten gibt es auch Erhaltungsmetadaten, die beschreiben, wie das digitale Objekt entstanden ist (z.B. wie eine Fotografie digitalisiert wurde, in welchem Zustand etc.).	RHM	Zeile 141-144

Kategorie	Bezeichnung der Kategorie	Definition	Quelle	Kriterien	Ankerbeispiel	Autor*in	Fundstelle
		ihre Versionen, ihre Verarbeitung, Speicherung und ihren Status beschreiben.					
5.4.	Der Einsatz von Recordkeeping-Metadaten für das Objekte-Management	Diese Arte der Metadaten wird dazu verwendet, die systematische Schaffung, Verwendung und Erhaltung, sowie die Zurverfügungstellung von digitalen Objekten festzuhalten. Dadurch sollen die Notwendigkeiten programmatischer, administrativer, finanzieller und rechtlicher Aspekte der jeweiligen Institution erfüllt werden.	vgl. NISO 2007:83		Zu den Recordkeeping-Metadaten kann ich nicht viel sagen, ich glaube die werden eher in einem anderen Umfeld verwendet (wie z.B. in Archiven). PHAIDRA beschäftigt sich mit digitalen Objekten und ich glaube, unter „record“ versteht man typischerweise etwas anderes, obwohl die Metadatenfelder, die man für deskriptive und obengenannte Metadatentypen braucht, wahrscheinlich großteils dieselben sind, die man beim Recordkeeping braucht.	RHM	Zeile 146-150

Kategorie	Bezeichnung der Kategorie	Definition	Quelle	Kriterien	Ankerbeispiel	Autor*in	Fundstelle
5.5.	Der Einsatz von strukturellen Metadaten	Strukturelle Metadaten werden verwendet, um die Elemente eines mehrteiligen digitalen Objekts miteinander in Beziehung zu setzen, oder um zusammengehörende Objekte in einem Paket zusammenzufassen. Wenn etwa die einzelnen Seiten eines Buches digitalisiert werden, können strukturelle Metadaten erfassen, in welche Reihenfolge die einzelnen Seiten des Buches gehören und wie sie mit der logischen Struktur, dem Inhaltsverzeichnis des Buches verbunden sind.	vgl. NISO 2007:83		Strukturelle Metadaten erlauben, dass die Bestände z.B. hierarchisch aufgebaut werden, oder dass manche Objekte enger verknüpft sind, oder man die Objekte versionieren kann, etc. Sie sind oft essenziell, um das Objekt in seinem Kontext zu verstehen.	RHM	Zeile 137-139
6.	Metadaten als digitale Objekte	Dieser Aspekt bezieht sich auf die Prinzipien für digitale Objekte der NISO. Diese sind im Wesentlichen in den bisherigen Kategorien enthalten. Ein wichtiger Punkt, der sich nicht in den Metadaten-Prinzipien findet, ist der Einsatz von persitenten Identifikatoren, wie er im der nächsten Kategorie dargestellt wird.	vgl. NISO 2007:85				

Kategorie	Bezeichnung der Kategorie	Definition	Quelle	Kriterien	Ankerbeispiel	Autor*in	Fundstelle
6.1.	Einsatz eines persistenten, eindeutigen Identifikators	Metadaten sollten mit einem persistenten, eindeutigen Identifikator (engl. persistent, globally unique identifier) versehen sein, der immer auf den Ort verweist, an dem sich die Metadaten zum jeweiligen Zeitpunkt befinden.	vgl. NISO 2007:52		Nein, die Metadaten selbst nicht, weil die sind veränderbar. Die Metadaten können die Owner jederzeit verändern. Das wird zwar mitdokumentiert, dass es verändert wurde. Aber man kann, weiß nicht, wenn ich jetzt draufkomme, ich habe vor fünf Jahren was reingestellt in PHAIDRA und habe dort einen Tippfehler oder ich habe eine zusätzliche Information, die ich unterbringen möchte, kann ich das jederzeit machen, also. Und dadurch, dass das verändert wird, wird das nicht viel Sinn machen. Und der DOI, der soll ja ohnehin nicht auf das Objekt selbst weisen, sondern auf eine Landing-Page. Und diese Landing-Page enthält ja die Metadaten, das heißt ich komme ja ohnehin auf die Metadaten	SB	#00:33:22#
7.	Die Modellierung von Metadaten: Metadatenkategorien und die Beziehungen	Die Modellierung von Metadaten findet sowohl auf der Ebene der Elemente, die Eingang in ein Metadaten-Modell (conceptual modeling) finden, als auch im Hinblick auf die Art, wie die Metada-	vgl. Zeng & Qin 2022:16	Diese Kategorie konzentriert sich auf das Conceptual Modeling und nicht auf die Darstellung der Metadaten. Als Bereich, für den modelliert wird, sollen	Die Fedora 3, auf der wir noch immer sind, war quasi XML-basiert und da hat das UW-Metadatenformat sehr gut reingepasst. In der Version 4 war es eine sogenannte Linked-Data-Plattform plötzlich. Und man hat RDF erwartet und / Man	RH	#00:06:17#

Kategorie	Bezeichnung der Kategorie	Definition	Quelle	Kriterien	Ankerbeispiel	Autor*in	Fundstelle
	zwischen ihnen	ten dargestellt werden (encoding modeling), statt. Beim Conceptual Modeling werden üblicherweise Metadaten-Elemente für einen bestimmten Bereich (eng. domain) definiert und eine Sammlung von Begriffen (engl. concepts) erstellt und deren strukturelle Beziehungen analysiert und festgehalten. Innerhalb des definierten Bereichs dienen diese Metadaten-Elemente und ihre Beziehungen dazu, Informationen zu managen, zu organisieren und aufzufinden.		hier die Bedarfe des gesamten digitalen Repositoriums und damit all seiner hochladenden Nutzer*innen gelten.	kann nach wie vor mit Metadaten als einem einfachen XML-File damit arbeiten. Machen auch sicher viele, weil dieser Umstieg auf RDF ist nicht trivial.		

Abkürzungsverzeichnis

BIBFRAME ..... *Bibliographic Framework Initiative*
BMWFW..... *Bundesministerium für Wissenschaft, Forschung und Wirtschaft*
DC ..... *Dublin Core*
DDC ..... *Dewey Decimal Classification*
IST..... *Institute for Science and Technology Austria*
JSON-LD ..... *JavaScript Object Notation for Linked Data*
KOS..... *knowlege organization system*
LAM..... *libraries, archives and museums*
LCSH ..... *Library of Congress Subject Headings*
LOD ..... *Linked Open Data*
LOM..... *Learning Object Metadata*
MODS ..... *Metadata Object Description Schema*
NISO ..... *National Information Standards Organization*
OAI-PMH..... *Open Archives Initiative Protocol for Metadata Harvesting*
ÖFOS ..... *Österreichische Systematik der Wissenschaftszweige*
PHAIDRA..... *Permanent Hosting, Archiving and Indexing of Digital Resources and Assets*
RDA ..... *Resource Description and Access*
RDF..... *Resource Description Framework*
RTF ..... *Rich Text Format*
SKOS ..... *Simple Knowledge Organization System*
UNVL ..... *University of Nevada, Las Vegas*
URI..... *Uniform Resource Identifier*
WCAG 2.0 ..... *Web Content Accessibility Guidelines 2.0*
XML..... *Extensible Markup Language*

Bibliographie

de Almeida Campos, Maria Luiza, Campos, Linair Maria & Barbosa, Nilson Theobald 2020. The Challenges of Semantic Interoperability in the Era of eScience on the Web. *Knowledge Organization* 47, 8, 680–695.

Archiv der Universität Wien 2019. *Nationale der Studentinnen der Medizinischen Fakultät Sommersemester 1907 A-Z*. https://phaidra.univie.ac.at/detail/o:944365 [Stand 2024-06-01].

Barroso, Isabel, Hartmann, Nadia & Ribeiro, Cristina 2015. Metadata Crosswalk for a Museum Collection in a Thematic Digital Library. *Journal of Library Metadata* 15, 1, 36–49.

Bellotto, Anna & Bettella, Cristiana 2019. Metadata as Semantic Palimpsests: The Case of PHAIDRA@unipd. In P. Manghi, L. Candela, & G. Silvello, hg. *Digital Libraries. Suppporting Open Science*. 15th Italian Research Conference on Digital Libraries, IRCDL 2019, Pisa, Italy, January 31 - February 1, 2019. Cham: Springer Nature Switzerland.

Biagetti, Maria Teresa 2021. Ontologies as Knowledge Organization Systems. *Knowledge Organization* 48, 2, 152–176.

Blumesberger, Susanne, Gründhammer, Veronika, Hudak, Rastislav, Schubert Bernhard & Traub, Imola Dora. 2016. *e-Infrastructures Austria. Deliverable. Cluster I: Metadatenkomplex aus nicht-technischer und technischer Sicht. Jobprofil von MetadatenmanagerInnen.* https://phaidra.univie.ac.at/o:441513 [Stand 2023-07-1].

Blumesberger, Susanne, Preza, José Luis & Zartl, Alexander. 2016. *e-Infrastructures Austria. Deliverable. Cluster I: Metadatenkomplex aus nicht-technischer und technischer Sicht. Umgang mit Metadaten in digitalen Repositorien.* https://phaidra.univie.ac.at/detail/o:441216 [Stand 2023-11-3].

Bogner, Alexander, Littig, Beate & Menz, Wolfgang 2014. *Interviews mit Experten: Eine praxisorientierte Einführung*. Wiesbaden: Springer Fachmedien.

Chan, Lois, Mai 2005. Metadata Interoperability: A Study of Methodology. http://www.white-clouds.com/iclc/cliej/cl19chan.htm [Stand 2022-10-7].

Chan, Lois, Mai & Zeng, Marcia Lei 2006a. Metadata Interoperability and Standardization. A Study of Methodology Part I. Achieving Interoperability at the Schema Level. *D-Lib Magazine* 12, 6. https://dlib.org/dlib/june06/chan/06chan.html [Stand 2022-05-9].

Chan, Lois, Mai & Zeng, Marcia Lei 2006b. Metadata Interoperability and Standardization - A Study of Methodology Part II. Achieving Interoperability at the Record and Repository Levels. *D-Lib Magazine* 12, 6. https://dlib.ejournal.ascc.net/dlib/june06/zeng/06zeng.html [Stand 2022-05-9].

Dahlberg, Ingetraud 1983. Conceptual Compatibility of Ordering Systems. *International Classification* 10, 5–8.

Dahlberg, Ingetraud 1981a. Conceptual Definitions for Interconcept. *International Classification* 8, 16–22.

Dahlberg, Ingetraud 1981b. Towards Establishment of Compatibility between Indexing Languages. *International Classification* 8, 2, 86–91.

Dresing, Thorsten & Pehl, Thorsten 2018. *Praxisbuch Interview, Transkription & Analyse. Anleitungen und Regelsysteme für qualitativ Forschende.* 8. Marburg: Eigenverlag.

EMMA 2024. EMMA - Educational Materials Made Accessible. Version 1.0—April 22, 2024. https://emma.lib.virginia.edu/doc/1.%20Introduction%20to%20EMMA%20%28Rev.%201.0.b1%29.pdf [Stand 2024-09-01].

FGDC 2000. Content Standard for Digital Geospatial Metadata Workbook (For use with FGDC-STD-001-1998). https://www.fgdc.gov/metadata/documents/workbook_0501_bmk.pdf [Stand 2024-06-1].

Fiala, Sonja 2019. Kennzeichnung barrierefreier Dateien - eine Zusammenstellung am Beispiel MARC21 und MODS. https://doi.org/10.25651/1.2019.0014.

Fiala, Sonja & Huggle, Christina 2019. Metadatenmapping: Die Gegenüberstellung verschiedener Metadatenschemata am Beispiel UWMETADATA>MODS 3.6. https://phaidra.univie.ac.at/o:1030317 [Stand 2021-03-15].

Gardiner, Alan Henderson 1947. *Ancient Egyptian Onomastica*. Oxford: Oxford University Press.

Gnoli, Claudio 2020. *Introduction to knowledge Organization*. London: Facet.

GO FAIR 2023a. *FAIR Principles*. FAIR Principles. https://www.go-fair.org/fair-principles/ [Stand 2023-08-18].

GO FAIR 2023b. *FAIR Principles. F1: (Meta)data are assigned globally unique and persistent identifiers.* FAIR Principles. https://www.go-fair.org/fair-principles/f1-meta-data-assigned-globally-unique-persistent-identifiers/ [Stand 2023-08-18].

GO FAIR 2023c. *FAIR Principles. I1: (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.* https://www.go-fair.org/fair-principles/i1-metadata-use-formal-accessible-shared-broadly-applicable-language-knowledge-representation/ [Stand 2023-08-18].

GO FAIR 2023d. *FAIR Principles. R1: (Meta)data are richly described with a plurality of accurate and relevant attributes.* FAIR Principles. https://www.go-fair.org/fair-principles/r1-metadata-richly-described-plurality-accurate-relevant-attributes/ [Stand 2023-08-18].

GO FAIR 2023e. *FAIR Principles. R1.3: (Meta)data meet domain-relevant community standards.* FAIR Principles. https://www.go-fair.org/fair-principles/r1-3-metadata-meet-domain-relevant-community-standards/ [Stand 2023-08-18].

Goody, Jack 1977. *The Domestication of the Savage Mind*. Cambridge: Cambridge University Press.

Gstrein, Silvia, Krexhammer, Andreas Preza, José Luis & Sakabe; Yukiko. 2016. *e-Infrastructures Austria. Deliverable. Cluster I: Metadatenkomplex aus nicht-technischer und technischer Sicht. Metadaten und Forschungsdaten.* https://phaidra.univie.ac.at/detail/o:441215 [Stand 2023-11-3].

Haselwanter, Thomas 2020. *e-Infrastructures Austria Plus. Projektbericht 2017-2019. Einleitung. Vorwort Haselwanter.* https://diglib.uibk.ac.at/4780259. [Stand 2024-01-06].

Hjørland, Birger 2008. What is Knowledge Organization (KO)? *Knowledge Organization* 35, 2/3, 86–101.

Hjørland, Birger & Gnoli, Claudio 2016. *Knowledge Organization*. https://www.isko.org/cyclo/knowledge_organization [Stand 2023-04-1].

Hodge, Gail 2001. *Metadata Made Simpler*. Bethesda, MD: NISO Press.

Hoffmann-Riem, Christa 1984. *Das adoptierte Kind. Familienleben mit doppelter Elternschaft.* München: Wilhelm Fink Verlag.

Hyman, Richard Joseph 1982. *Shelf Access in Libraries*. Chicago: American Library Association.

Jeitler, Andreas & Blumesberger, Susanne. 2016. *e-Infrastructures Austria. Deliverable. Cluster I: Metadatenkomplex aus nicht-technischer und technischer Sicht. Metadata and Accessibility.* https://phaidra.univie.ac.at/detail/o:459817 [Stand 2023-11-3].

Kallmeyer, Werner & Schütze, Fritz 1976. Konversationsanalyse. *Studium Linguistik* 1, 1–28.

Kempf, Andreas Oskar u. a. 2014. New Ways of Mapping Knowledge Organization Systems: Using a Semi-Automatic Matching Procedure for Building up Vocabulary Crosswalks. *Knowledge Organization* 41, 1, 66–75.

Kopácsi, Sandor, Hudak, Rastislav & Ganguly, Raman 2017. Implementation of A Classification Server to Support Metadata Organization for long-term Preservation Systems. *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen & Bibliothekare* 70, 2, 225–243.

Kopácsi, Sandor, Hudak, Rastislav Preza & José Luis. 2016. *e-Infrastructures Austria. Deliverable. Cluster I: Metadatenkomplex aus nicht-technischer und technischer Sicht. Classification Server*.
https://phaidra.univie.ac.at/o:441217 [Stand 2023-07-1].

Krexhammer, Andreas, Preza, José Luis & Sakabe, Yukiko 2016. *e-Infrastructures Austria. Deliverable. Cluster I: Metadatenkomplex aus nicht-technischer und technischer Sicht. Use of Metadata om Research Data Repositories: Use Case IST Austria DataRep*.
https://phaidra.univie.ac.at/detail/o:441219 [Stand 2023-11-3].

Kuckartz, Udo, Dresing Thorsten, Radiker Stefan & Stefer, Claus. 2008. *Qualitative Evaluation. Der Einstieg in die Praxis*. Wiesbaden: Verlag für Sozialwissenschaften.

Lamnek, Siegfried 2005. *Qualitative Sozialforschung: Lehrbuch*. Weinheim: Beltz/Juventa.

Lauruhn, Michael & Groth, Paul 2016. Sources of Change for Modern Knowledge Organization Systems. *Knowledge Organization* 43, 8, 622–629.

Maisel, Thomas 2013. *Archiv der Universität Wien, Bildersammlung*.
https://phaidra.univie.ac.at/detail/o:289959 [Stand 2023-08-20].

Marcondes, Carlos Henrique 2013. Knowledge Organization and Representation in Digital Environments: Relations Between Ontology and Knowledge Organization. *Knowledge Organization* 40, 2, 115–122.

Mayr, Philipp & Petras, Vivien 2008. Building a Terminology Network for Search: The KoMoHe Project. In J. Greenberg & W. Klas, hg. *Metadata for Semantic and Social Applications: Proceedings of the International Conference on Dublin Core and Metadata Applications: Berlin, 22-26 September 2008: DC 2008: Berlin, Germany.* Göttingen: Universitätsverlag Göttingen, 177–182.

Mayring, Philipp 2022. *Qualitative Inhaltsanalyse. Grundlagen und Techniken.* 13., überarbeitete Auflage. Weinheim Basel: Beltz.

Mayring, Philipp 1983. *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. Weinheim Basel: Beltz.

Neville, Hugh 1970. Feasibility Study of a Scheme for Reconciling Thesauri Covering a Common Subject. *Journal of Documentation* 26, 4, 313–336.

Neville, Hugh H. 1972. Thesaurus Reconciliation. *Aslib Proceedings* 24, 620–626.

NISO 2007. A Framework of Guidance for Building Good Digital Collections https://www.niso.org/sites/default/files/2017-08/framework3.pdf [Stand 2024-01-06].

NISO 2023. *All About NISO*. https://www.niso.org/what-we-do [Stand 2023-04-1].

Preza, José Luis 2016. *e-Infrastructures Austria. Deliverable. Cluster I: Metadatenkomplex aus nicht-technischer und technischer Sicht. Metadata Quality.* https://phaidra.univie.ac.at/detail/o:441218 [Stand 2023-11-3].

Ranganathan, Shiyali Ramamrita 1967. *Prolegomena to Library Classification*. 3. Auflage New York: Asia Publishing House.

Riley, Jenn 2004. Understanding Metadata. https://groups.niso.org/higherlogic/ws/public/download/17446/Understanding%20Metadata.pdf [Stand 2023-04-2].

Schütze, Fritz 1992. Sozialarbeit als „bescheidene“ Profession. In B. Dewe, W. Ferchhoff, & F.-O. Radtke, hg. *Erziehen als Profession. Zur Logik professionellen Handelns in pädagogischen Feldern*. Opladen: Leske + Budrich.

Southwick, Silvia B. 2015. A Guide for Transforming Digital Collections Metadata into Linked Data Using Open Source Technologies. *Journal of Library Metadata* 15, 1, 1–35.

Subirats, Imma, Solodovnik, Iryna, Budroni Paolo, Ganguly Raman & Hudak, Rastislav. 2017. Metadata and Controlled Vocabularies Alignment in PHAIDRA international persistent Ecosystem. *Information Services & Use* 37, 1, 25–32.

Universität Wien 2023a. *e-Infrastructures Austria*. https://e-infrastructures.univie.ac.at/home/index.htm [Stand 2023-07-2].

Universität Wien 2023b. *Über PHAIDRA -Services*. Repositorienmanagement PHAIDRA-Services. https://datamanagement.univie.ac.at/ueber-phaidra-services/ [Stand 2023-08-18].

Universitätsbibliothek Universität Wien 2023. *Phaidra - Permanent Hosting, Archiving and Indexing of Digital Resources and Assets*. https://bibliothek.univie.ac.at/phaidra.html [Stand 2023-08-25].

W3C 2024. *Resource Description Framework (RDF)*. https://www.w3.org/RDF/ [Stand 2024-05-11].

W3C 2023. *XML Essentials - W3C*. https://www.w3.org/standards/xml/core [Stand 2023-04-1].

Wilkinson, Mark D., Dumontier, Michel, Aalbersberg, IJsbrand Jan, Appleton, Gabrielle, Axton, Myles, Baak, Arie, Blomberg, Niklas, Boiten, Jan u. a. 2016. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data* 3, 1, 160018. https://doi.org/10.1038/sdata.2016.18.

Zeng, Marcia Lei 2008. Knowledge Organization Systems. *Knowledge Organization* 35, 160–182.

Zeng, Marcia Lei & Qin, Jian 2022. *Metadata*. Third edition. London: Facet.