



Modellierung und Dokumentation von Use-Cases für wissenschaftliche Sammlungen z.B. in METS, EAD, TEI-HEADER, LIDO, PREMIS und DCCD (R 4.2.1)

Version 20. Mai 2015
Cluster 4
Verantwortlicher Partner HAB

DARIAH-DE Aufbau von Forschungsinfrastrukturen für die e-Humanities

Dieses Forschungs- und Entwicklungsprojekt wird / wurde mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBF), Förderkennzeichen 01UG1110A bis N, gefördert und vom Projektträger im Deutschen Zentrum für Luft- und Raumfahrt (PT-DLR) betreut.

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Projekt: DARIAH-DE: Aufbau von Forschungsinfrastrukturen für die e-Humanities

BMBF Förderkennzeichen: 01UG1110A bis N

Laufzeit: März 2011 bis Februar 2016

Dokumentstatus: Final

Verfügbarkeit: öffentlich

Autoren: Peter Andorfer, David Wolf (HAB)

Revisionsverlauf:

| Datum | Autor | Kommentare |
|------------|----------------|--------------------------|
| 07.04.2015 | Peter Andorfer | Einleitung, Gliederung |
| 27.04.2015 | David Wolf | Anmerkungen, Korrekturen |
| 07.05.2015 | David Wolf | Kapitel 2.5 PREMIS |
| 11.05.2015 | Peter Andorfer | Kapitel 2 und 3 |
| 15.05.2015 | David Wolf | Anmerkungen, Korrekturen |

Inhaltsverzeichnis:

| | |
|--|----------|
| 1. Einleitung | 3 |
| 1.1. Gliederung | 5 |
| 2. Uses-Cases (wissenschaftlicher Sammlungen) | 5 |
| 2.1. TEI-Header] | 5 |
| 2.1.1. Use-Case: Beschreibung einer Sammelhandschrift | 6 |
| 2.1.2. Uses-Case: Beschreibung eines linguistischen Textkorpus | 7 |
| 2.1.3. Möglichkeiten und Grenzen | 8 |
| 2.2. EAD | 8 |
| 2.2.1. Use-Case: Ein Findbuch | 9 |
| 2.2.2. Möglichkeiten und Grenzen | 10 |
| 2.3. METS | 11 |
| 2.3.1. Use-Case: Digitale Edition | 12 |
| 2.3.2. Use-Case: DFG –Viewer | 13 |
| 2.3.3. Möglichkeiten und Grenzen | 13 |
| 2.4. LIDO | 13 |
| 2.4.1. Möglichkeiten und Grenzen | 16 |

| | |
|---|-----------|
| 2.5. PREMIS..... | 17 |
| 2.5.1. Use-Case:..... | 18 |
| 2.5.2. Möglichkeiten und Grenzen | 18 |
| 2.6. DCCAP | 19 |
| 2.6.1. Use-Case: DFG Praxisregeln „Digitalisierung“ | 20 |
| 2.6.2. Use-Case: DARIAH-DE Collection Application Profiles (DCAP) | 20 |
| 2.6.3. Möglichkeiten und Grenzen | 21 |
| 3. Fazit | 22 |
| 3.1. Ausblick..... | 23 |
| 4. Literaturverzeichnis | 24 |
| 5. Abbildungsverzeichnis | 25 |

1. Einleitung

Der Titel dieses Reports suggeriert, dass es ebenso einen klar definierten Begriff „wissenschaftliche Sammlungen“ gibt wie eine gelebte Praxis, solch wissenschaftliche Sammlungen unter Verwendung von wenigstens sechs verschiedenen Metadatenstandards in maschinenlesbarer Form zu beschreiben. Dabei zeigt aber bereits die Auflösung der einzelnen Akronyme, dass die wenigsten dieser Standards primär für die Beschreibung wissenschaftlicher Sammlungen konzipiert wurden. Dies trifft besonders auf den „Encoded Archival Description“-Standard (EAD) zur Beschreibung von Findbüchern in Archiven zu, wie auch für die Richtlinien der Text Encoding Initiative (TEI), welche dazu dienen, Text in all seinen Spielarten und Facetten systematisch zu erfassen. LIDO bzw. „Lightweight Information Describing Objects“ wiederum stammt aus dem Museumskontext und rückt vornehmlich die einzelnen Objekte einer Sammlung als die Sammlung selbst in den Vordergrund. Das Objekt steht auch im Zentrum von PREMIS „Preservation Metadata Implementation Strategies“, wobei Objekt bzw. Object hier definiert wird als: „discrete unit of information in digital form“.¹

Im Gegensatz dazu ist der „**Metadata Encoding & Transmission Standard**“ auf keine spezifische Objektgruppe ausgerichtet, es sei denn, man fasst Metadaten selbst als eine eigene Objektgruppe auf. Ähnlich wie bei PREMIS dient METS aber zur Beschreibung digitaler Objekte.

Für die Beschreibung von Sammlungen wurden somit nur die „**Dublin Core Collection Description Terms**“² konzipiert, welche im Titel dieses Reports unter dem Akronym DCCD zusammengefasst werden. Im Unterschied zu den übrigen Akronymen steht hinter DCCD aber kein eigenständiger Metadatenstandard inklusive ausführlicher Dokumentation oder eigenem XML-Schema, sondern eine Sammlung von „new properties, classes, vocabulary encoding schemes and syntax encoding schemes, created for use by the application profile for collection-level description developed by the Dublin Core Collection Description Task Group.“³ Das hier erwähnte (**Dublin Core**)

¹ PREMIS Editorial Committee, *PREMIS Data Dictionary for Preservation Metadata*, 2012, 5.

² <http://dublincore.org/groups/collections/collection-terms/>

³ Ebd.

Application Profile for Collection-Level Description wird im weiteren Text als Dublin Core Collection Application Profile bezeichnet bzw. als DCCAP.⁴

Ebensowenig wie einen allgemein anerkannten und weit verbreiteten Metadatenstandard zur Sammlungsbeschreibung, gibt es auch eine allgemein anerkannte und weit verbreitete Definition von wissenschaftlichen Sammlungen. Die Erarbeitung und Diskussion einer solchen Definition ist jedoch eines der Ziele des „Stakeholdergremiums Wissenschaftliche Sammlungen“, welches sich im Rahmen der bisherigen Treffen auf folgenden Kriterienkatalog verständigen konnte.

Eine „Sammlung wissenschaftlich relevanter digitaler Forschungsdaten

- besteht aus diskreten, voneinander verschiedenen, logisch unabhängigen Einheiten und wurde intentional angelegt (**Bilderegel**);
- muss nicht aus einem Forschungsanliegen heraus entstanden sein, ist aber Gegenstand wissenschaftlicher Fragestellungen bzw. hat das Potential zu einer wissenschaftlichen Analyse und dient der Validierung von Aussagen, Methoden, Thesen, Hypothesen oder Theorien in Forschung und Lehre (**Reproduzierbarkeit**);
- kann sowohl Ursprung als auch Ergebnis wissenschaftlicher Arbeit sein (**Research Data Life Cycle**);
- ist in einer regelhaften Form maschinenlesbar, dokumentiert, idealerweise nach internationalen Standards erfasst und mit Normdaten ausgezeichnet (**Prozessier- und Interpretierbarkeit**);
- gibt Auskunft über ihren Rechtsstatus (z.B. **Nutzungsbedingungen / Lizenzen**);
- dient der Ordnung der Sammlungsgegenstände und der archivischen Sicherung (**Archivierung**).“⁵

Während die Summe dieser Kriterien ein sehr klar umrissenes Bild einer wissenschaftlichen Sammlung zeichnet, wird im weiteren Textverlauf aber bewusst ein deutlich weniger scharf konturierter Sammlungs-begriff verwendet. Dies hat den Grund, dass dadurch eine größere Zahl potentieller Use-Cases von Sammlungen generiert werden kann und somit Grenzen und Möglichkeiten der einzelnen Metadatenstandards im Kontext von Modellierung und Dokumentation von

⁴ <http://dublincore.org/groups/collections/collection-application-profile/>

⁵ Protokoll Treffen Stakeholdergremium Wissenschaftliche Sammlungen in Göttingen am 17. Oktober 2014, S. 4 [eigene Zählung, Original unpaginiert]; online [nicht öffentlich] unter: <https://dev2.dariah.eu/wiki/download/attachments/31429961/Ergebnisprotokoll%20Stakeholdergremium.pdf?version=1&modificationDate=1418900169201&api=v2>. „Digitale Forschungsdaten“ werden im gleichen Dokument definiert als: „Daten, unabhängig ihrer Provenienz, werden im Kontext einer geistes- und kulturwissenschaftlichen Forschungsfrage zu digitalen Forschungsdaten, sobald sie gesammelt, beschrieben, ausgewertet und / oder erzeugt und in maschinenlesbarer Form zum Zwecke der Nachvollziehbarkeit von Forschungsergebnissen sowie zur Archivierung, Zitierbarkeit und Weiterverarbeitung aufbewahrt werden. Eine Aggregation von Forschungsdaten (n>1) ist in diesem Sinne eine Wissenschaftliche Sammlung.“

(wissenschaftlichen) Sammlungen besser oder überhaupt erst sichtbar gemacht werden können.

1.1. Gliederung

Im weiteren Verlauf des vorliegenden Reports, genauer in Kapitel 2, geht es daher darum, eine Reihe konkreter Szenarien zu skizzieren, in denen Sammlungen im Kontext geisteswissenschaftlicher Projekte bereits in standardisierter und maschinenlesbarer Form modelliert und dokumentiert wurden. Dabei basiert die Zusammenstellung der einzelnen hier vorgestellten Szenarien eben auf einer sehr weitgefassten Definition von (wissenschaftlicher) Sammlung. Als Sammlung werden hier nämlich intendiert zusammengeführte, diskrete, physische oder analoge Objekte verstanden. Kapitel 3 fasst die Eckpunkte und Ergebnisse des vorigen Kapitels zusammen und formuliert Vorschläge für weitere Arbeitsschritte.

2. Use-Cases (wissenschaftlicher Sammlungen)

Ein Großteil der vorzustellenden Use-Cases stammt aus dem Umfeld der Herzog August Bibliothek Wolfenbüttel. Als ausgewiesene Forschungsbibliothek begegnet man hier nämlich dem Begriff der Sammlung einerseits aus der Perspektive einer Gedächtnisorganisation mit einem genau vorgegebenen Sammlungsauftrag, andererseits aber auch aus fachwissenschaftlicher Forschungsperspektive. Darüber hinaus ist die Herzog August Bibliothek auch DARIAH-DE Projektpartner und an den innerhalb DARIAH-DEs entwickelten Konzepten zum Thema wissenschaftliche Sammlungen sowie Möglichkeiten und Praktiken hinsichtlich Dokumentation und Modellierung derselben maßgeblich interessiert.

Bei der Auswahl der weiteren Use-Cases spielte hingegen das Kriterium der Verfügbarkeit von konkretem Datenmaterial die ausschlaggebende Rolle.

2.1. TEI-Header

Dass ein Standard, der von einer Einrichtung namens „Text Encoding Initiative“ betreut wird, zur Beschreibung von Text konzipiert wurde, sollte keine Überraschung darstellen. Wenn im Kontext von Sammlungsmodellierung und -dokumentation also auf TEI zurückgegriffen wird, so liegt der Schluss nahe, dass es sich bei den Objekten der zu beschreibenden Sammlung auch um Texte handelt. Will man dann auch noch eine Verbindung zwischen Text und Sammlung herstellen, so bieten sich, gerade in einem bibliothekarischen Umfeld, die bibliothekarischen Einheiten Sammelband, -werk und -ausgabe sowie Sammelhandschrift an, die alle mehrere distinkte Texte enthalten, wobei bei einem Sammelwerk die Zusammenstellung ausgewählter Texte ein neues und urheberrechtliches Werk konstituiert.⁶

⁶ Zu den konkreten Unterschieden zwischen Sammelband, -werk, -ausgabe und -handschrift siehe Severin Corsten and Claus W. Gerhardt, eds., *Lexikon Des Gesamten Buchwesens. Bd. 6: Phraseologie - Schütz, 2.*, völlig neubearb. Aufl. (Stuttgart: Hiersemann, 2003), 475–478.

2.1.1. Use-Case: Beschreibung einer Sammelhandschrift

Standardmäßig erfolgt die systematische, menschen- wie maschinenlesbare Beschreibung von (Sammel-) Handschriften derzeit mittels TEI-P5, den „Richtlinien für die Auszeichnung und den Austausch elektronischer Texte“. Darin wird die Handschrift bzw. die der TEI-Datei zugrundeliegende Handschrift, im Element <msDesc> (/TEI/teiHeader/fileDesc/sourceDesc/msDesc) als ein eigenständiges Objekt beschrieben: „(manuscript description) contains a description of a *single identifiable* manuscript or other text-bearing object.“⁷ Dieses Element kann mehrere spezifische Kindelemente – teilweise auch mehrmals – beinhalten (<additional>, <history>, <msContents>, <msIdentifier>, <msPart>, <physDesc>). Außerdem ist innerhalb von <msDesc> auch die Verwendung anderer Element-Gruppen wie vor allem etwa der „core“-Gruppe zulässig. Dies bedeutet, dass mit Hilfe der Elemente und Attribute des TEI-Header und deren vorgegebener Reihenfolge und Struktur eine detaillierte Beschreibung der Handschrift bzw. der Sammlung selbst möglich ist.⁸ Eine Minimalbeschreibung einer Handschrift (<msDesc>) sollte entsprechend den DFG-Praxisregeln „Digitalisierung“ jedoch mindestens die Elemente <head>/<title>, <msIdentifier>/<settlement>, <msIdentifier>/<repository> und <msIdentifier>/<idno> verwenden.⁹

Die einzelnen Texte der Handschrift bzw. die Objekte der Sammlung werden wiederum in jeweils eigenen <msPart>-Elementen beschrieben, welche laut TEI-P5 Informationen „about an *originally distinct* manuscript or part of a manuscript“ enthalten.¹⁰

Zusätzlich zu den Angaben zur Sammelhandschrift und den darin anzutreffenden Texten sind im TEI-Header auch noch weitere Elemente zur Beschreibung der Handschrift respektive zum entsprechenden XML-TEI Datensatz verpflichtend erforderlich. Dazu zählt vor allem das komplexe Element namens „title statement“ (<titleStmnt>). Darin werden neben anderen Informationen auch die für die Erstellung der Beschreibung verantwortlichen Personen sowie der Zeitpunkt, wann die Beschreibung angelegt wurde, dokumentiert.

Ein konkretes Beispiel einer Beschreibung einer Sammelhandschrift in TEI kann über den Webauftritt der Herzog August Bibliothek abgerufen werden.¹¹

⁷ <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-sourceDesc.html>, eigene Hervorhebung durch Kursivsetzung.

⁸ Für eine vollständige Auflistung sämtlicher zulässiger Elemente (und Attribute) siehe die TEI-P5 Spezifikation, online unter www.tei-c.org.

⁹ DFG - Deutsche Forschungsgemeinschaft, “DFG-Praxisregeln ‘Digitalisierung,’” February 2013, 3f, http://www.dfg.de/formulare/12_151/12_151_de.pdf. Dort können auch knappe Erläuterungen zu den genannten Elementen gefunden werden. Für Empfehlungen einer ausführlicheren Beschreibung siehe Sebastian Meyer et al., *TEI-Anwendungsprofil Für Digitalisierte Handschriften*, 2014.

¹⁰ <http://www.tei-c.org/release/doc/tei-p5-doc/de/html/ref-msPart.html>, eigene Hervorhebung durch Kursivsetzung.

¹¹ <http://diglib.hab.de/?db=mss&list=ms&id=12-helmst&catalog=Haertel-Lesser&mode=xml>.

2.1.2. Use-Case: Beschreibung eines linguistischen Textkorpus

Neben den im vorigen Abschnitt behandelten Textsammlungen aus dem Umfeld des Buch- und Bibliothekswesens begegnen Forscherinnen und Forscher Sammlungen von Texten vor allem auch in linguistischen Korpora. Das „Metzler Lexikon Sprache“ definiert den Begriff Korpus in einem engeren Sinne explizit als „*Sammlung* einer möglichst hohen, notwendigerweise aber immer begrenzten Anzahl möglichst zusammenhängender sprachlicher Äußerungen (gesprochen oder/geschrieben) [...]“¹²

Aktuell wird dabei zur systematischen und maschinenlesbaren Beschreibung von Korpora vermehrt auf die Richtlinien der TEI zurückgegriffen. So auch im „Long-term Access and Usage of Deeply Annotated InformaTION“ oder auch LAUDATIO-Repository“, einem von der Deutschen Forschungsgemeinschaft (DFG) finanziertem Repository historischer Korpora. Derzeit umfasst das Repository – bei dem es sich strenggenommen ebenfalls um eine Sammlung handelt – 13 verschiedene Korpora. Jedes dieser Korpora ist mit Hilfe eines auf die speziellen Bedürfnisse eines historischen Korpus angepassten Sets von Metadatenfeldern beschrieben, wobei Namen, Typen, Attribute, Anzahl und Anordnung dieser Felder weitgehend auf jenen des TEI-Headers basiert.

Aus einer solchen Beschreibung können somit Informationen zum Korpus (zur Sammlung) selbst entnommen werden, etwa der Name des Korpus (/teiCorpus/teiHeader/fileDesc/titleStmt/title), der Name des Herausgebers (/teiCorpus/teiHeader/fileDesc/titleStmt/editor/persName), Informationen über rechtliche Aspekte (/teiCorpus/teiHeader/fileDesc/publicationStmt/availability) oder einer knappen, menschenlesbaren Korpusbeschreibung (/teiCorpus/teiHeader/encodingDesc/projectDesc).

Gleichzeitig enthält die Beschreibung aber auch Informationen zu den den Korpus konstituierenden Texten (Objekten), und zwar sowohl in summarischer als auch in text- oder objektspezifischer Form. Zu ersterer zählen Angaben wie zum Gesamtumfang des Korpus (/teiCorpus/teiHeader/fileDesc/extent), welcher hier als Summe aller Tokens (@type=Tokens“) angegeben wird, oder zu den in den Texten verwendeten Sprachen (/teiCorpus/teiHeader/profileDesc/langUsage/language). Die objektspezifischen Angaben sind neuerlich in jeweils eigenen <teiHeader>-Elementen (/teiCorpus/teiCorpus/teiHeader) notiert, die sich durch das Attribut-Value-Pair @type=DocumentHeader vom <teiHeader>-Element des gesamten Korpus unterscheiden. Das <teiHeader>-Element des Korpus (/teiCorpus/teiHeader) hat das Type-Attribut „CorpusHeader“. Hinzu kommt noch ein weiteres <teiHeader>-Element (/teiCorpus/teiCorpus/teiHeader), welches mit dem Type-Attribut „PreperationHeader“ versehen ist, worin Informationen zur Vorbereitung der Annotation der Texte dokumentiert sind.¹³ Die Annotationen selbst, die Deklaration des Tagsets geschieht wiederum in Reihe verschiedener Elemente wie <tagsDecl> oder

¹² Helmut Glück, ed., *Metzler-Lexikon Sprache*, 2., überarb. und erw. Aufl. (Stuttgart [u.a.]: Metzler, 2000), 384. Eigene Hervorhebung durch Kursivsetzung.

¹³ Siehe dazu auch die LAUDATIO-Repository Documentation, online unter: <http://www.laudatio-repository.org/repository/documentation>.

<tagUsage>, die in der Sektion <encodingDesc> (/teiCorpus/teiHeader/encodingDesc) versammelt sind.

Eine konkrete Beschreibung eines im LAUDATIO-Repositorys gespeicherten mit Hilfe von TEI-Header Elementen beschriebenen Korpus kann über den Webauftritt von LAUDATIO eingesehen werden.¹⁴

2.1.3. Möglichkeiten und Grenzen

Die beiden Use-Cases zeigen, dass Sammlungen von Texten unter Verwendung der Elemente des TEI-Headers umfangreich und detailliert beschrieben werden können. Dies trifft sowohl auf die Gesamtheit der Texte, also auf die Sammlung als Ganzes, als auch auf die einzelnen Texte, die Objekte der Sammlung zu. Darüber hinaus können im TEI-Header auch einige essentielle Informationen zur Sammlungsbeschreibung respektive zur TEI-Datei selbst kodiert werden. Dazu zählen etwa Angaben zu den für die Erstellung der Sammlungsbeschreibung verantwortlichen Personen und Organisationen, Auskünfte zum Entstehungsdatum der Beschreibung sowie zu allfälligen Änderungen und Revisionen, wie auch generelle Lizenzangaben. In solche TEI-Sammlungsbeschreibung können außerdem auch Informationen über den größeren Bezugsrahmen der Textsammlung kodiert werden. So wird im Element <msIdentifier> (/TEI/teiHeader/fileDesc/sourceDesc/msDesc/msIdentifier) über die Kindelemente <repository> und <collection> etwa der Bestand identifiziert, dem die Sammelhandschrift zugeordnet ist.

Auch wenn es angesichts des Namens des Metadatenstandards *Text* Encoding Initiative naheliegend ist, so muss doch betont werden, dass die TEI-Header Elemente primär für die Erfassung und Beschreibung von Texten konzipiert wurden. Gleichzeitig gilt es aber zu beachten, dass Text stets an ein gewisses Trägermedium gebunden ist, dessen Eigenschaften mithilfe des TEI-Vokabulars ebenfalls dokumentiert werden kann. Dies trifft auf genuin naheliegende Materialien wie Papier oder Pergament ebenso zu, wie beispielsweise auch auf Stein.¹⁵

2.2. EAD

Bei EAD oder **E**lectronic **A**rchival **D**escription handelt es sich um einen XML-Standard zur Kodierung von Findbüchern, worunter Inventare, Indizes oder Repertorien verstanden werden, die zur Orientierung in Archivbeständen angelegt wurden. Herausgegeben und betreut wird der Standard von der Library of Congress.¹⁶ Mittels EAD werden somit strenggenommen nicht die eigentlichen Beständen (oder Sammlungen) eines Archivs beschrieben, sondern die dazu angefertigten Findbehelfe. Dass durch eine Beschreibung von Findbüchern oder Inventaren indirekt auch eine Beschreibung der darin verzeichneten Objekte bzw. des erfassten Bestandes erfolgt,

¹⁴ <http://hdl.handle.net/11022/0000-0000-1F5B-9>.

¹⁵ Siehe dazu etwa das Projekt Epidoc <http://sourceforge.net/p/epidoc/wiki/Home/>.

¹⁶ Vgl. <http://www.loc.gov/ead/eadabout.html>.

ist jedoch unbestritten. Insofern handelt es sich bei einem nach EAD kodierten archivalischem Repertorium um eine standardisierte und menschen- wie maschinenlesbare Sammlungsbeschreibung.¹⁷

2.2.1. Use-Case: Ein Findbuch

Trotz der derzeitigen systematischen Kodierung von Findbüchern nach EAD in vielen Archiven Deutschlands und den Bemühungen, die entsprechenden Ergebnisse niederschwellig online zugänglich zu machen, konnten keine hier verwertbaren EAD-Dateien aus einem Deutschen Archiv gefunden werden. Stellvertretend dafür muss hier daher auf eine Beispieldatei zurückgegriffen werden, die auf der Website eaddiva.com veröffentlicht wurde¹⁸. Dies stellt jedoch insofern kein Problem dar, da es sich bei EAD ja ohnehin um eine internationale Empfehlung zur standardisierten Beschreibung von Findbüchern handelt.

Eine EAD-Datei bzw. deren Wurzel-Element <ead> beinhaltet stets die beiden Elemente <eadheader> und <archdesc>. Während im <eadheader>-Element Informationen zum jeweiligen Findbehelf kodiert werden, enthält das <archdesc>-Element Angaben zu dem durch den Findbehelf erschlossenen Bestand als Gesamtes, wie ggf. auch zu einzelnen Untergruppen und Objekten, die den Bestand ausmachen.¹⁹

Im Kontext von Sammlungen würde dies bedeuten, der <eadheader> identifiziert die Sammlungsbeschreibung (<eadid>) und enthält bibliographische Angaben über Verfasser, Titel, Entstehungszeit und -kontext (<filedesc>), sowie Auskünfte über Kodierungsrichtlinien (<proposedesc>) und Änderungen in der Sammlungsbeschreibung (<revisiondesc>). Allerdings sind nur die beiden erstgenannten Elemente verpflichtend.²⁰

Inhaltliche, strukturelle, administrative und kontextuelle Informationen zur Sammlung wie auch zu ihren Objekten befänden sich somit hingegen allesamt im <archdesc>-Element.

Hierin ist aber nur das <did>-Element verpflichtend, wobei <did> für Descriptive Identification steht und als „Hüllenelement“ fungiert, „das andere Elemente mit den Kerninformationen über die erschlossenen Materialien [...] zusammenfasst.“²¹ Zu diesen Kerninformationen gehören unter anderen Angaben zur Provenienz

¹⁷ Angelika Menne-Haritz and Michael J. Fox, trans., *Das EAD-Kochbuch. (ohne Abschnitte 4 Und 5) Übersetzung Des EAD Cook Book (Vers. 1) von Michael J. Fox, 2003*, <http://www.bundesarchiv.de/imperia/md/content/daofind/eadkochbuch.pdf>.

¹⁸ <http://eaddiva.com/sampleEAD/pitt-mss297.xml>.

¹⁹ Vgl. RLG, *RLG Best Practice Guidelines for Encoded Archival Description* (Mountain View, California, 2002), 2, <http://www.oclc.org/content/dam/research/activities/ead/bpg.pdf?urlm=161431>. Anke Löbnitz and Angelika Menne-Haritz, *Encoded Archival Description Tag-Library. Version 2002, Übersetzt von Anke Löbnitz Und Angelika Menne-Haritz* (Berlin, 2006), 115 und 47, <http://www.bundesarchiv.de/imperia/md/content/daofind/1.pdf>.

²⁰ Löbnitz and Menne-Haritz, *Encoded Archival Description Tag-Library. Version 2002, Übersetzt von Anke Löbnitz Und Angelika Menne-Haritz*, 115.

²¹ *Ibid.*, 100.

(<origination>) und zum Titel des Bestandes (<unittitle>), zu dessen äußerer Erscheinung (<physdesc>), eine kurze menschenlesbare Beschreibung (<abstract>) oder die Nennung des Archivs, das den Bestand beheimatet (<repository>). Die einzelnen Elemente selbst verfügen wiederum über mehrere Kindelemente, was eine detaillierte und strukturierte (Sammlungs-)Beschreibung ermöglicht.

In <bioghist> (**B**iography or **H**istory) einem optionalen Kindelement von <archdesc> kann der historische Kontext des Bestandes dargelegt werden, sowohl in Form eines kurzen Textes oder einer strukturierten Zeitleiste.

<scopecontent>, ein nächstes Kindelement von <archdesc> beschreibt den Umfang und die Gegenstände des Materials sowie damit in Verbindung stehende relevante Organisationen, Ereignisse oder Personen mit dem Ziel, potentiellen Nutzern zu helfen die Relevanz des Bestandes für die eigene Recherche einschätzen zu können. Ein Ziel, dass von einem weiteren <archdesc>-Kindelement unterstützt wird, nämlich von <controlaccess>, worin der Bestand mit Hilfe von kontrolliertem Vokabular standardisiert charakterisiert wird.

Angaben über Nutzungsmöglichkeiten und -rechte enthalten die Elemente <accessrestrict> und <userrestrict>, Vorschläge für eine bevorzugte Zitierweise des Bestandes enthält hingegen das Element <prefercite>.

Was die einzelnen Objekte der Sammlung bzw. die Teilbestände anbelangt, so werden diese im optionalen <dsc>-Element (**D**escription of **S**ubordinate **C**omponents) beschrieben, wobei darin auch die Beziehung der einzelnen Teilbestände zueinander bzw. deren hierarchische Gruppierung dokumentiert ist. Die Entstehungs- oder Entwicklungsgeschichte der im <dsc>-Element abgebildeten Struktur wiederum beschreiben die Elemente <processinfo> (**P**rocessing **I**nformation), <acqinfo> (**A**cquisition **I**nformation) und <accruals>, natürlich mit entsprechender inhaltlicher Differenzierung.

2.2.2. Möglichkeiten und Grenzen

Die Möglichkeiten zur Modellierung und Dokumentation von Sammlungen mittels EAD sind zweifelsfrei vorhanden. Dies trifft ebenso auf Sammlungs- und Objektebene zu, welche im <archdesc>-Element zu erfolgen hätte, als auch auf die Ebene der Sammlungsbeschreibung, wofür sich die Kindelemente des <eadheader>-Elements anbieten würden. Außerdem gibt es kaum objektbedingte Einschränkungen hinsichtlich der Verwendung von EAD, im Unterschied etwa zum zuvor beschriebenen TEI-Header welcher primär für die Beschreibung von Text verwendet wird.

Gegen eine Verwendung von EAD zur Sammlungsmodellierung und -dokumentation sprechen, abgesehen von dem generellen Einwand, dass es sich EAD strenggenommen um einen Standard zur Beschreibung von Findbehelfen handelt, einmal der pragmatische Grund des damit potentiell einhergehenden Arbeitsaufwandes, als auch die prinzipielle Frage nach den Unterschieden zwischen Sammlung und Archivbestand.

Zu ersterem: EAD definiert 146 Elemente und rund 70 Attribute und stellt somit ein umfangreiches und mächtiges Vokabular zur Verfügung, mit dessen Hilfe ebenso

umfangreiche Beschreibungen von Sammlungen angelegt werden könnten. Gleichzeitig zeigt sich EAD sehr großzügig, was die Anzahl der verpflichtenden Angaben betrifft und schreibt nur wenige Pflichtfelder vor. Daraus folgt, dass Umfang und Detailgrad von Sammlungsbeschreibungen in EAD sehr stark variieren können, wodurch das Nutzungspotential von solchen Beschreibungen konsequenterweise aber eingeschränkt wird.

Zu zweitem: Eine Stärke von EAD liegt mit Sicherheit darin, die oftmals komplexe Binnengliederung eines in der Regel historisch gewachsenen Archivbestandes abbilden zu können. Gerade darin kann aber ein systematischer Unterschied zu einer Sammlung erkannt werden, nämlich dann, wenn man eine Sammlung primär nur über die Beziehung Sammlung-Objekt definiert und allfällige Objekt-Objekt-Beziehungen unbeachtet lässt. Zu fragen wäre außerdem, ob Teilbestände nicht als eigene Sammlungen zu beschreiben wären.

Zu guter Letzt sei auch noch darauf hingewiesen, dass EAD außerhalb der Archiv-Szene weniger bekannt sein dürfte etwa zuvor gesprochene der TEI-Header.

2.3. METS

Während die Fachgebiete, Disziplinen und Personengruppen die mit TEI und EAD arbeiten sich an verschiedenen Stellen wieder mit Disziplinen der Geisteswissenschaften überschneiden, dürfte die Schnittmenge an Geisteswissenschaftlern die sich mit METS auseinandersetzen doch deutlich kleiner ausfallen. Dies ist allerdings nur eine persönliche Einschätzung, da allfällige Untersuchungen zum allgemeinen Bekanntheitsgrad der hier zu behandelnden Metadatenstandards innerhalb der Geisteswissenschaften nicht bekannt sind. Ein Grund für den (behaupteten) geringeren Bekanntheitsgrad von METS (**M**etadata **E**ncoding & **T**ransmission **S**tandard) dürfte in dessen vornehmlich auf digitale Objekte konzentrierten Anwendungsfällen liegen. Denn wo mit TEI, wie auch mit EAD großteils reale und physische Objekte wie Handschriften, Briefe, Tagebücher oder Akten, Urkunden, und Findbücher beschrieben werden, steht bei METS vor allem die Verwaltung und Strukturierung von digitalen Objekten bzw. Dateien im Vordergrund.²² Eine Verwaltung, die noch dazu über den reinen Archivierungsaspekt hinausgeht und die Realisierung konkreter Anwendungen in den Vordergrund rückt. Anwendungen beispielsweise, welche eine kontextualisierte und interaktive Anzeige einzelner Objekte im Webbrowser ermöglichen.

Den Anforderung solcher Anwendungen versucht METS gerecht zu werden, indem es die einzelnen Elemente in insgesamt sieben Bereiche unterteilt, wovon aber nur der Bereich `<structMap>` oder **Structural Map** verpflichtend ist.²³

`<structMap>` beschreibt die (logische oder physische) Struktur eines Objektes, wobei sich im Falle einer Sammlung, die mit Hilfe von METS beschrieben wird, die Frage

²² Digital Library Federation, *<METS> Metadata Encoding and Transmission Standard: Primer and Reference Manual*, 2010, 15, <http://www.loc.gov/standards/mets/METSPrimerRevised.pdf>.

²³ Vgl. dazu wie in weiterer Folge Digital Library Federation, *<METS> Metadata Encoding and Transmission Standard: Primer and Reference Manual*.

stellt, ob bei einer Sammlung überhaupt irgend eine Art von Struktur i.e. eine Reihenfolge der Objekte der Sammlung sinnvoll beschrieben werden kann oder soll. Innerhalb des <structMap>-Elements bzw. innerhalb darin enthaltener **divisions** bzw. <div>-Elemente, wird mittels **File-Pointer-Element** <fptr> respektive dem Wert des darin enthaltenen FILEID-Attributs die Verknüpfung zu den entsprechenden digitalen Objekten bzw. deren Metadaten realisiert, die im Bereich **file Section** <fileSec> zu beschreiben sind.

Die konkreten Dateien werden in <fileSec>, im Element <Flocat> oder **File Location**, ein Kindelement eines <file>-Elements referenziert und zwar mittels URL, URN, PURL oder HANDLE als Wert eines xlink:href-Attributs von <Flocat>. Zusätzlich dazu können die jeweiligen <file>-Elemente noch in verschiedene **File Groups** <FileGrp>-Elementen organisiert werden.

Keiner der eben beschriebenen Bereiche enthält jedoch Informationen über das hier in METS zu beschreibende Objekt. Für inhaltliche oder beschreibende Metadaten wie beispielsweise Entstehungszeit oder Erzeuger, ist der Bereich <dmdSec> bzw. **Descriptive Metadata Section** vorgesehen. Hier können die entsprechenden Infos entweder direkt in die METS-Datei kodiert werden, oder man referenziert eine externe Datei. Entscheidend ist aber, dass die jeweiligen Metadaten in verschiedenen Metadatenstandards wie etwa MODS oder TEI eingebunden werden können. Insofern verlagert sich zu einem gewissen Grad die Frage, ob und wie gut METS zur Modellierung und Dokumentation von Sammlungen geeignet ist, auf die einzubindenden Standards.

Für rechtliche oder administrative Aspekte zu dem in METS beschriebenen Objekt ist wiederum der Bereich <amdSec> oder **Administrative Metadata Section** zuständig. Analog zu <dmdSec>-Sektion können auch die Inhalte von <amdSec> in die METS-Datei kodiert oder externe Datenquellen referenziert werden.

Informationen über die METS-Datei, etwa wann und von wem sie angelegt wurde, wann sie verändert wurde und zu welchem Zwecke werden, analog zu TEI oder EAD in einem Header-Element, dem <METSHdr> notiert.

Die beiden noch nicht besprochenen Bereiche <behaviorSec> und <structLink> bzw. **Behavior Section** und **Structural Linking Section** wiederum sind vor allem für die jeweiligen Anwendungsszenarien für welche die METS-Datei angelegt wurde relevant, weshalb sie hier nicht weiter vorgestellt werden sollen. Vorgestellt werden sollen hier hingegen zwei konkrete Anwendungsgebiete, in denen mittels METS-Dateien das Zusammenspiel verschiedener, mehr oder weniger diskreter Objekte organisiert wird.

2.3.1. Use-Case: Digitale Edition

Digitale Editionen können dann als wissenschaftliche Sammlung verstanden werden, sofern man sich darauf verständigen möchte, die einzelnen Bestandteile einer solchen Edition als distinkte Objekte anzuerkennen, welche auch jenseits des spezifischen Kontextes der digitalen Edition sinnvoll existieren könnten. Realistischerweise ist dies jedoch nicht bei allen Objekten einer solchen 'Sammlung' in letzter Konsequenz argumentierbar. Digitale Editionen, die etwa in der Reihe „Editiones Electronicae Guelferbytae“ veröffentlicht wurden, bestehen (meist) aus einer oder mehreren TEI-

Dateien, welche eine annotierte Transkription der zu edierenden Texte enthalten, Faksimiles der edierten Texte, TEI-Dateien für diverse Register und Bibliographien, einer ebenfalls in TEI kodierte Einleitung sowie Stylesheets und Skripten, für die Anzeige-, Navigations- und Suchfunktionen der Edition. Diese Dateien bilden die Objekte der Sammlung, wohl wissend jedoch, dass ein aus diesem Kontext herausgelöstes Personenregister beispielsweise, nur bedingt sinnvoll verwendet werden kann.

Ein Beispiel einer METS-Datei einer in der Reihe „Editiones Electronicae Guelferbytanae“ veröffentlichten digitalen Edition kann im Rahmen des Webauftrittes der Wolfenbütteler Digitalen Bibliothek (WDB) eingesehen werden.²⁴

2.3.2. Use-Case: DFG –Viewer

Ein verbreitetes Anwendungsszenario für METS-Dateien ist die Implementierung des DFG-Viewers bzw. die Aufbereitung und Beschreibung einzelner Bilddateien für eine Darstellung derselben im DFG-Viewer. Im Regelfall handelt es sich bei diesen Bilddateien um Faksimiles von Drucken und Handschriften. Im Gegensatz zur digitalen Edition ist die Bezeichnung eines Digitalisates eines alten Druckes als Sammlung kaum gerechtfertigt, da die einzelnen Seiten eines Buches bzw. in Analogie dazu die Digitalisate der einzelnen Seiten nur schwerlich als distinkte Objekte verstanden werden können, die auch losgelöst ihres eigentlichen Kontextes sinnvoll verwendbar sind.

2.3.3. Möglichkeiten und Grenzen

Die große Stärke von METS liegt in der Beschreibung von Strukturen und Abhängigkeiten digitaler Objekte bzw. derer Metadaten vor dem Hintergrund konkreter Anwendungsszenarien. Dies bedeutet umgekehrt jedoch, dass METS einen Großteil seiner Vorzüge gegenüber anderen Metadatenstandards einbüßt, sobald die METS-Datei aus dem Kontext der Anwendung genommen wird. Wenn es also 'nur' darum geht, eine Sammlung vorwiegend mit Hilfe inhaltlicher wie administrativer Metadaten zu modellieren und zu dokumentieren, so bedarf es dazu in der Regel keines „Metadata Encoding & Transfer Standards.“ Dies zeigt sich alleine schon darin, dass METS ja andere Metadatenstandards gerade für die Dokumentation von administrativen und inhaltlichen Informationen einbindet.

Gegen ein Verwendung von METS zur Modellierung und Dokumentation von Sammlungen spricht vor allem aber auch – dies sei der Vollständigkeit halber noch einmal wiederholt – dass METS für digitale Objekte konzipiert wurde.

2.4. LIDO

LIDO is a schema intended for delivering metadata, for use in a variety of online services, from an organization´s online service, from an organization´s online col-

²⁴ <http://diglib.hab.de/edoc/ed000145/mets.xml>

lections database to portals of aggregated resources, as well as exposing, sharing and connecting data on the web. The strength of LIDO lies in its ability to support the full range of *descriptive information about museum objects*.²⁵

Der für den Museumskontext entwickelte Metadatenstandard **L**ightweight **I**nformation **D**escribing **O**bjects (LIDO) beschreibt museale Objekte. Obwohl LIDO den Objektbegriff aber bereits im Namen trägt konnte in den eingesehen einschlägigen LIDO-Materialien²⁶ keine grundlegende Definition dazu gefunden werden. Dies ist insofern auch nachvollziehbar, da der Begriff des musealen Objektes wohl weitgehend selbsterklärend sein dürfte. Dies wird etwa aus den in den LIDO-Materialien verwendeten Beispielen deutlich, handelt es sich dabei etwa um ein Gemälde, einen Brunnen,²⁷ bzw. um ein Mineral, eine Kragenflasche, einen Holzschnitt, eine Fotografie und um einen Schlosspark.²⁸

Das Fehlen einer wenigstens rudimentären Definition wird somit erst aus der Perspektive eines Sammlungsbegriffes evident, welcher sich als Summe wenigstens zweier distinkter Objekte versteht. Die Frage, die sich bei einer Auseinandersetzung mit LIDO aus Perspektive einer Sammlung also aufdrängt, lautet, ob der von LIDO verwendete Objektbegriff auch eine Sammlung als Objekt anerkennt. Dies scheint, zumindest theoretisch durchaus der Fall zu sein. Wirft man nämlich einen Blick in die Dokumentation des LIDO-Standards bzw. in das XML-Schema, so findet sich bei der Beschreibung des Elements <lido:category> die ausdrückliche Empfehlung sich bei der Kategorisierung des zu beschreibenden Objektes an den „CIDOC-CRM concept definitions“ zu orientieren. Unter den anschließenden Beispielen, die zeigen sollen wie das Element <lido:category> konkret zu befüllen wäre, befindet neben den Konzepten „Man-Made Object“ (E22) und „Man-Made Feature“ (E25) auch das Konzept E78 „Collection“. Die Autoren der CIDOC-CRM Spezifikation beschreiben dabei „Collection“ wie folgt:

This class [collection, Anm. PA] comprises aggregations of instances of E18 Physical Thing that are assembled and maintained (“curated” and “preserved,” in museological terminology) by one or more instances of E39 Actor over time for a specific purpose and audience, and according to a particular collection development plan.²⁹

Es darf also davon ausgegangen werden, dass zu den mit LIDO beschreibbaren Objekten auch Sammlungen zählen. Außerdem darf wohl auch davon ausgegangen werden, dass Objektbeschreibungen in LIDO sehr detailliert und umfangreich ausfallen können. Denn einerseits sollen die Objektbeschreibungen in möglichst vielen

²⁵ ATHENA Project, *Lightweigh Information Describing Objects (LIDO): The International Harvesting Standard for Museums* (Rom, 2011), 22, <http://www.athenaeurope.org/getFile.php?id=786>, eigene Hervorhebung durch Kursivsetzung.

²⁶ ATHENA Project, *Lightweigh Information Describing Objects (LIDO): The International Harvesting Standard for Museums*. Erin Cobum et al., *LIDO - Lightweight Information Describing Objects Version 1.0*, 2010. “What Is LIDO,” accessed February 22, 2013, <http://network.icom.museum/cidoc/working-groups/data-harvesting-and-interchange/what-is-lido/>.

²⁷ <http://network.icom.museum/cidoc/working-groups/lido/lido-overview/examples/>.

²⁸ <http://www.lido-schema.org/documents/LIDO-Introduction.pdf>.

²⁹ Nick Crofts et al., *Definition of the CIDOC Conceptual Reference Model*, 2009, <http://www.cidoc-crm.org/crm-concepts/>. Siehe hier “E78 Collection”.

Applikationen („variety of online services“) Verwendung finden und andererseits sollen die Beschreibungen aus qualitativ wie quantitativ sehr unterschiedlichen Quellen gespeist werden können.

Die einzelnen Elemente zur Objektbeschreibung teilen sich in LIDO auf sieben verschiedene Gruppen bzw. „information groups“ auf, welche selbst wiederum Teil des Wrapper-Elementes `<descriptiveMetadata>` oder `<administrativeMetadata>` sind. Daneben gibt es außerdem noch das bereits erwähnte `<category>`-Element.

Die erwartungsgemäß umfangreichere Gruppe bilden die Elemente der **beschreibenden Metadaten** (`<administrativeMetadata>`), deren Elemente ihrerseits wieder den Wrapper-Elementen `<objectClassificationWrap>`, `<objectIdentificationWrap>`, `<eventWrap>` und `<objectRelationWrap>` eingeschrieben sind, wobei jedoch nur die Elemente `<objectClassificationWrap>` und `<objectIdentificationWrap>` verpflichtend zu verwenden sind.

Wie aus den Elementnamen schon hervorgeht, werden in `<objectClassificationWrap>` Informationen über den Typ des Objektes dokumentiert. Im Falle einer Sammlungsbeschreibung wäre hier als Typ wohl Sammlung. In diesem Element können aber außerdem noch weitere klassifizierende Angaben gemacht werden, etwa zum Stil, zum Alter oder zur wissenschaftlichen Fachrichtung der Sammlung.

Im Element `<objectIdentificationWrap>` hingegen werden all jene Daten gesammelt, die zur Identifizierung des Objektes beitragen. Dazu zählen neben Informationen zum Titel des Objektes `<titleWrap>` und einer Beschreibung `<objectDescriptionWrap>` auch Angaben zum Standort `<repositoryWrap>` oder zur Größe des Objektes `<objectMeasurementsWrap>`. Ein Gutteil der für eine Sammlungsbeschreibung wesentlichen Informationen kann somit dem Element `<objectIdentificationWrap>` entnommen werden.

Über die fakultative Klasse `<eventWrap>` werden Beziehungen der Ressource zu 'events' beschrieben wobei unter Events hier Ereignisse wie beispielsweise die Aufnahme einer Sammlungstätigkeit oder die Produktion der gesammelten Gegenstände verstanden werden. In der LIDO-Spezifikation wird das `<event>`-Element folgendermaßen beschrieben: „Identifying, descriptive and indexing information for the events in which the object participated or was present at, e.g. creation, excavation, collection, and use.“³⁰ Über die verschiedenen Kindelemente von `<event>` werden außerdem räumliche `<eventPlace>`, zeitliche `<eventDate>` und personelle wie institutionelle `<eventActor>` Bezüge zum Objekt bzw. zu dem jeweils mit dem Objekt verbundenen Event hergestellt.

Das ebenfalls optionale Element `<objectRelationWrapper>` ist ein: „Wrapper for information about related topics and works, collections, ect“.³¹ Inhaltlich-thematische Informationen zum Objekt werden im Kindelement `<subjectWrap>` notiert und

³⁰ Erin Cobum et al., *LIDO - Lightweight Information Describing Objects Version 1.0*, 2010, 46.

³¹ Ibid., 76.

Verweise auf verwandte Ressourcen, Werke, Sammlungen u.ä. im Kindelement <relatedWorksWrap> festgehalten.

Die administrativen Metadaten im Wrapper-Element <**administrativeMetadata**> beinhalten (1.) rechtliche Informationen zum Objekt <**rightsWorkWrap**>, (2.) Informationen über bereits vorhandene (Meta)daten, Katalogeinträge oder sonstige Aufzeichnungen („records“) zum Objekt <**recordWrap**> und (3.) Angaben über etwaige (digitale) Ressourcen, Stellvertreter oder Derivate („surrogates“) des Objektes <**resourceWrap**> wie etwa Digitalisate.³² Verpflichtend notwendig ist jedoch nur das Element <recordWrap>.

2.4.1. Möglichkeiten und Grenzen

Das systematische Beschreiben von Objekten respektive Sammlungen ist mittels LIDO sowohl umfassend als auch detailliert möglich. Durch die Integration von CIDOC-CRM ist außerdem ein hohes Maß an Interoperabilität gewährleistet. Mit diesen Vorzügen geht jedoch auch ein relativ großer Arbeitsaufwand bei der Erstellung einer Sammlungsbeschreibung einher. Will man in LIDO etwa das Entstehungsdatum eines Objektes kodieren, so bedarf dafür folgender Struktur:

```
6 <lido:eventWrap>
7   <lido:eventSet>
8     <lido:event>
9       <lido:eventType>
10        <lido:term>Herstellung</lido:term>
11      </lido:eventType>
12 > <lido:eventActor> [20 lines]
33 </lido:eventActor>
34 <lido:eventDate>
35   <lido:date>
36     <lido:earliestDate>1654</lido:earliestDate>
37     <lido:latestDate>1654</lido:latestDate>
38   </lido:date>
39 </lido:eventDate>
40 </lido:event>
41 </lido:eventSet>
42 </lido:eventWrap>
```

Abbildung 1: Entstehungsdatum eines Objektes in LIDO.

Die LIDO-Datei des Brunnens „Fontana del Moro“ am Piazza Navona in Rom bzw. einer Reihe von Fotografien des Brunnens aus dem Bestand des Bildarchivs Foto Marburg ist rund 1300 Zeilen lang.³³ Zum Vergleich, die METS-Datei zur oben vorgestellten digitalen Editionen ist nur etwa halb so lange und die TEI-Datei der ebenfalls bereits vorgestellten Beschreibung einer Sammelhandschrift umfasst 905

³² Ibid., 87, 100, 103.

³³ Bei dieser Datei handelt es sich um eine der offiziellen LIDO-Beispieldateien, die auf der LIDO-Website heruntergeladen werden können: http://www.lido-schema.org/documents/examples/LIDO-Example_FMobj20344012-Fontana_del_Moro.xml. Der Codeausschnitt in Abbildung 1 stammt aus dieser Datei.

Zeilen, wobei darin auch Informationen zur Materialbeschaffenheit einzelner Seiten oder Transkriptionen von Incipits enthalten sind.

Ob sich der Aufwand der Erstellung einer LIDO-Datei zum Zwecke von Sammlungsbeschreibungen lohnt, hängt schlussendlich jedoch vom konkreten Anwendungsfall ab.

2.5. PREMIS

Bei PREMIS („**PRE**servation **M**etadata: **I**mplementation **S**trategies“) handelt es sich in erster Linie um eine 2003 gegründete Arbeitsgruppe, die 2008 ein Datenlexikon und Informationen zur Langzeitarchivierung veröffentlichte.³⁴ Ein Schema zur Darstellung von PREMIS in XML wird ähnlich wie EAD seitdem von der Library of Congress unterhalten und weiterentwickelt.³⁵ Nach Eigenaussage steht PREMIS „als Teilmenge im Zentrum“ aller Metadaten, „die eine Organisation benötigt, die ein digitales Langzeitarchiv betreibt.“³⁶

Das PREMIS-Datenlexikon basiert auf dem Referenzmodell ISO 14721 des **O**pen **A**rchival **I**nformation **S**ystem (OAIS), um alternative Strategien zu Speicherung, Kodierung und Metadatenmanagement zu untersuchen und zu evaluieren.³⁷

Bei der Beschreibung im Datenlexikon liegt der Fokus darauf, wie Informationen gespeichert sind. PREMIS beschreibt „semantische Entitäten“ anstelle von „Metadatenelementen“ aus dem Ziel heraus, eine Interoperabilität zwischen Repositorien und maschinelle Lesbarkeit von Informationen zu ermöglichen. Die fünf Entitäten sind intellektuelle Entitäten, Objekte, Ereignisse, Rechte und Agenten, deren Beziehungen untereinander beschrieben werden.³⁸ Die *Objekte* sind unterteilt in die Untertypen *file*, *bitstream* und *representation* und definiert als „discrete unit of information in digital form“.³⁹

Bewusst wird auf die in Bibliotheken gebräuchliche Definition, nach der ein Objekt aus Identifier, Metadaten und Daten besteht, verzichtet, um Attribute und ihre Beziehungen darzustellen. Je nach Anforderungen können große Mengen struktureller Metadaten in Formaten wie beispielsweise METS oder MPEG-21⁴⁰ erforderlich werden. Diese Metadatensätze bilden in PREMIS eigene Objekte. Die einzige verpflichtende semantische Einheit ist <objectIdentifier>.

³⁴ Vgl. aktuelle Version aus dem Jahr 2012: PREMIS Editorial Committee, *PREMIS Data Dictionary for Preservation Metadata*.

³⁵ Vgl. <http://www.loc.gov/standards/premis/>.

³⁶ Priscilla Caplan, *PREMIS Verstehen*, trans. Tobias Beinert, 2009, 5, http://www.loc.gov/standards/premis/understanding_premis_german.pdf.

³⁷ The Consultative Committee for Space Data Systems (CCSDS), *Reference Model for an Open Archival Information System (OAIS)*, Recommended Practices (Washington D.C., 2012), <http://public.ccsds.org/publications/archive/650x0m2.pdf>.

³⁸ PREMIS Editorial Committee, *Introduction and Supporting Materials from PREMIS Data Dictionary for Preservation Metadata*, 2012, 6, <http://www.loc.gov/standards/premis/v2/premis-report-2-2.pdf>.

³⁹ Ibid.

⁴⁰ Vgl. <http://mpeg.chiariglione.org/standards/mpeg-21>

2.5.1. Use-Case:

Aufgrund fehlender Originaldaten wird in diesem Fall erneut auf ein Beispiel zurückgegriffen, das in der Dokumentation der **Library of Congress (LoC)** beschrieben ist.⁴¹

Es handelt sich um ein TIFF-Bild mit den *handles* <objectIdentifier> und <relatedObjectIdentification>. Das Bild ist im Archiv über den eindeutigen Identifikator <objectIdentifier> auffindbar, aber der Ursprungsname wurde geändert (<originalName> „001h.tif“). Die Kategorie <objectCategory> gibt an, dass es eine Datei ist, deren Prüfsumme <fixity> im MD5-Format von der der LoC berechnet wurde. Zwei Arten von Formatinformation werden gegeben, MIME-Typ und Version <formatDesignation>, sowie der <formatRegistryName> PRONOM. Ebenfalls erwähnt werden die bei der Erstellung eingesetzten Anwendungen, das Scanprogramm <creatingApplication> ScandAll und <creatingApplication> Adobe Photoshop. <nhibtors> gibt an, dass die Datei unverschlüsselt ist, über <contentLocation> wird der Ort der Datei angegeben.

Die Besonderheiten von PREMIS kommen bei der Genauigkeit, mit der Betriebssystem <environment>, Festplattensystem <software> und technische Mindestanforderungen <storageMedium> des Rechners <hardware> angegeben werden, deutlich zum Vorschein. Eine Reproduktion oder Migration der Datei wird durch die hier gespeicherten Angaben deutlich erleichtert. Auch die Verortung der TIFF-Datei und ihre Beziehung zu anderen Teilen einer größeren Repräsentation <relationship> sind dokumentiert. Da diese Geschwister <sibling> sind, können wir schließen, dass es sich ebenfalls um TIFFs handelt.

2.5.2. Möglichkeiten und Grenzen

Sammlungsmodellierung mit PREMIS ist technisch prinzipiell möglich, doch es gibt besser geeignete, weil anders fokussierte Modelle. Während PREMIS auf Objektebene feine Granularität aufweist (95 Semantische Einheiten für *Objekte*, 19 für *Ereignisse*, 13 für *Agenten*, 68 für *Rechte*), werden auch die ausgelassenen Bereiche deutlich. Besonders die *intellektuellen Entitäten* müssen mit Verweis auf deskriptive Metadaten an anderer Stelle beschrieben sein.⁴²

PREMIS ist zwar als internationale Richtlinie und Standard gedacht, aber zur Versammlung von für die Langzeitarchivierung wichtigen Metadaten. Das wird bereits durch die Selbstdarstellung als „Teilmenge“ überdeutlich. Sich ändernde Bezüge innerhalb von Sammlungen, die sich eventuell auch noch regelmäßig dynamisch ändern, können nicht einfach wiedergegeben werden. Beschreibt man nur die Beziehungen der Objekte untereinander, wäre eine Sammlung ebenfalls nur ein weiteres Objekt.

⁴¹ Caplan, *PREMIS Verstehen*.

⁴² PREMIS Editorial Committee, *Introduction and Supporting Materials from PREMIS Data Dictionary for Preservation Metadata*, 23.

2.6. DCCAP

Die Verfasser der Dokumentation des **Dublin Core Collections Application Profile** (DCCAP), die Mitglieder der Dublin Core Collection Description Task Group weisen gleich zu Beginn ihres Dokumentes darauf hin, dass es sich dabei nicht um die Beschreibung eines XML-Formats handeln würde: „This document is **not** a description of an XML format“ um aber gleich im nächsten Satz darauf aufmerksam zu machen, dass mehrere Möglichkeiten der Serialisierung eines solchen Dublin Core Application Profiles (DCAP) nach XML und anderen Auszeichnungsformaten denkbar sind: „There may be multiple bindings of this DCAP, to XML and to other syntaxes.“⁴³

Unter einem Application Profile bzw. einem Dublin Core Application Profile wiederum verstehen die Autoren die Auswahl von Beschreibungsfeldern („set of terms“) aus dem Dublin Core Namensraum sowie Richtlinien zu deren Verwendung, Gruppierung, Anordnung, Häufigkeit und zu den erlaubten oder möglichen Inhalten und Werten der jeweiligen Beschreibungsfelder. Außerdem wird in einem Application Profile auch festgelegt, was überhaupt beschrieben wird. Im Falle eines Dublin Core **Collection** Application Profiles handelt es sich dabei konsequenterweise natürlich um eine Sammlung. Wie jedoch ebenfalls explizit hervorgehoben wird, zählen dazu aber auch Kataloge und Indizes, sprich Aggregationen von Metadaten, die eine Sammlung beschreiben („This documents describes a DC Application Profile for describing a collection, a catalogue or index, i.e. an aggregation of metadata that describes a collection.“).

Konkret beschreibt das Application Profile die sechs verschiedenen Entitäten („entity types“) Collection, Item, Location, Service, Catalog or Index und Agent und deren Beziehungen zueinander. Eine Sammlung (Collection) ist etwa durch einen Katalog erschlossen, ein Agent kann eine Sammlung besitzen, ebenso wie er auch einen Service administriert, ein Service erlaubt den Zugriff auf eine Sammlung, etc.

Um die Besonderheit einer Sammlung und das Zusammenspielen der einzelnen Entitäten auch dokumentieren zu können, hat Dublin Core zusätzlich zu den bereits bestehenden Dublin Core Elementen und Terms im Namensraum cld ein Set von „Collection Description Terms“ definiert. Diese Collection Description Terms wiederum bestehen aus fünf Elemente mit deren Hilfe die Eigenschaften <itemType>, <itemFormat>, <dateItemsCreated>, <isLocatedAt>, <isAccessedVia>, <catalogueOrIndex>, <associatedCollection> und <describedCollection> beschrieben werden können. Hinzu kommen noch Elemente zur Definition von Encoding-Schemes für Angaben zu Bereichen wie <AccrualMethod>, <Frequency> oder <AccrualPolicy>.⁴⁴

⁴³ Hier wie im Folgenden siehe: *Dublin Core Collections Application Profile* (Dublin Core Collection Description Task Group, March 9, 2007), <http://dublincore.org/groups/collections/collection-application-profile/>.

⁴⁴ Aufgrund der weitgehend für sich selbst sprechenden Namen, wird von einer näheren Erklärung der Elemente Abstand genommen. Eine solche kann unter nachfolgendem Link eingesehen werden: <http://purl.org/cld/terms/>.

2.6.1. Use-Case: DFG Praxisregeln „Digitalisierung“

Beispiele zweier konkreter Serialisierungen des DCCAP können den DFG-Praxisregeln „Digitalisierung“ entnommen werden.⁴⁵ Es handelt sich dabei einmal um die Beschreibung einer Sammlung von Abzügen und Glasnegativen, so der Inhalt der <cld:itemType>-Element aus den Jahren 1869-1900 (<cld:dateItemsCreated>), die von Ludwig Bickell, (<dc:creator>) dem Begründer des Marburger Universitätsmuseums angelegt wurde und inhaltlich um die Kunst, Architektur und Hessen (<dc:subject>) kreist.

Das zweite Beispiel beschreibt die „Virtuelle Sammlung“ (<dc:type>) „Helmstedter Drucke Online“, die an der Herzog August Bibliothek Wolfenbüttel verortet ist (<cld:isLocatedAt>) und über eine Website bzw. URL abgerufen werden kann (<cld:isAccessedVia>). Bei den Objekten der Sammlung handelt es sich um Images und Fulltext <cld:itemType>, die zwischen 2010 und 2014 erstellt wurden <cld:dateItemsCreated>.

Dass bei diesen beiden Beispielen keinerlei Angaben über die Sammlungsbeschreibung gemacht wurden, etwa wann oder durch wen die Beschreibung angelegt wurde, entspricht dabei den Empfehlungen der DFG. Allerdings sehen diese Empfehlungen auch weder Angaben über den Entstehungszeitraum der Sammlung selbst noch über die Objekte der Sammlung vor. Interessanterweise verweist die DFG hinsichtlich der Auswahl der Beschreibungsfelder auf den ISO-Standard 27730 „Information and Documentation – International Standard Collection Identifier“.⁴⁶ Wie zu sehen war, finden sich in beiden Beispielen jedoch sehr wohl Datierungen zu den Sammlungsobjekten – selbst wenn dies offenbar den DFG-Praxisregeln widerspricht.

2.6.2. Use-Case: DARIAH-DE Collection Application Profiles (DCAP)

Die Ausarbeitung einer auf dem DCCAP basierenden Spezifizierung eines Datenmodells zur Sammlungsbeschreibung wurde in der ersten Förderphase von DARIAH-DE in der „[AG Daten & Sammlungen](#)“⁴⁷ begonnen. Eine Publikation allfälliger Ergebnisse ist nicht erfolgt, alternativ dazu liegt jedoch eine umfassende, jedoch nur DARIAH-DE intern einsehbare Wiki-Seite vor.⁴⁸

Ein Vergleich des **DARIAH-DE Collection Application Profiles (DCAP)** mit dem DCCAP zeigt eine Reduktion von sechs auf vier Klassen oder Entitäten. So beschreibt das DCAP nur noch die Klassen Collection, Agent, Service und Location, die beiden weiteren DCCAP-Klassen Item und Catalogue or Index werden nicht berücksichtigt. Die Gründe für diese Entscheidung werden nicht thematisiert.

⁴⁵ DFG - Deutsche Forschungsgemeinschaft, „DFG-Praxisregeln ‘Digitalisierung.’“

⁴⁶ Da dieser neun Seiten umfassende Standard hinter einer Paywall von 58 CHR verborgen liegt, wird dieser 'Standard' in diesem Paper nicht weiter berücksichtigt.
http://www.iso.org/iso/catalogue_detail.htm?csnumber=44293

⁴⁷ <https://dev2.dariah.eu/wiki/x/jAYj>

⁴⁸ <https://dev2.dariah.eu/wiki/x/VoFn>

Abgesehen von den bekannten Elementen aus dem Dublin Core Namensraum werden im DCAP noch eine Reihe weiterer Elemente eingeführt, die unter dem selbst definierten Namensraum „dclap“ zusammengefasst sind, wobei **dcalp** vermutlich für **D**ariah **C**ollection **L**evel **A**pplication **P**rofiles steht. Diese Elemente dienen vornehmlich für die Kodierung technischer Informationen der Collection- und der Service-Klasse. So enthält <dclap:metadataEncodingScheme>, als einem Kindelement der Collection-Klasse Informationen über ein allfälliges Schema, „welches für die Metadaten der Objekte der Sammlung verwendet wurde“. <dclap:itemEncodingScheme>, ebenfalls ein Kindelement der Collection-Klasse wiederum bezeichnet das Schema, „welches in den Objekten der Sammlung verwendet wird“. Gemäß DCAP werden diese Informationen als Freitext angeführt. Die Verwendung dieser Elemente ist darüber hinaus optional.

Von den insgesamt 13 unterschiedlichen Elementen der Service-Klasse stammen gleich neun aus dem dclap-Namensraum. Von diesen sind die vier Elemente <dclap:serviceAccessMethod>, <dclap:serviceFunction>, <dclap:serviceURL> und <dclap:accessControl> verpflichtend anzugeben. Optional sind hingegen die Elemente <dclap:acronym>, <dclap:subset>, <dclap:schemaID>, <dclap:serviceInterface> und <dclap:serviceHelpURL>. Abgesehen vom Element <dclap:serviceURL>, dessen Inhalt vom Typ URL sein muss, ist für die Inhalte der übrigen verpflichtenden Elemente ein jeweils spezifisches kontrolliertes Vokabular vorgesehen. Um welche Werte es sich dabei aber konkret handelt, kann nur aus der DARIAH-DE Collection-Registry (CR) bzw. aus deren graphischer Oberfläche, der CR-GUI in Erfahrung gebracht werden.

Gerade am Beispiel der Service-Klasse respektive aufgrund der darin enthaltenen und eigens definierten Elemente wird jedoch deutlich, dass das DCAP für digitale Sammlungen bzw. Sammlung digitaler Objekte konzipiert wurde, wobei diese Sammlung im besten Falle auch noch über eine OAI-PMH Schnittstelle verfügen sollte. Dies ist insofern nachvollziehbar, da das DCAP vor dem Hintergrund der DARIAH-DE Collection Registry entwickelt wurde.

Das DCAP liegt derzeit in drei konkreten Serialisierungen vor, nämlich einmal als XML-Schema, einmal in Form der CR-GUI bzw. in den dafür entworfenen Java-Klassen und einmal in den durch OAI-PMH exportierten Datensätzen. Prekärerweise stimmen die drei Serialisierungen erstens untereinander nicht zur Gänze überein und entsprechen zweitens auch nicht zur Gänze dem DCAP.⁴⁹

2.6.3. Möglichkeiten und Grenzen

Der größte Vorteil einer auf dem DCCAP basierenden Sammlungsbeschreibung basiert in erster Linie wohl darauf, hier auf ein extra für die Beschreibung von Sammlungen konzipiertes Vokabular zurückgreifen zu können. Ein weiterer Vorteil liegt außerdem in den im DCCAP konzipierten sechs Klassen und der Beschreibung ihrer Beziehungen zueinander. So lassen sich nämlich nicht nur komplexere Sammlungsstrukturen und -hierarchien abbilden (Sammlung A ist Teil einer Sammlung

⁴⁹

Vgl.

ausführlicher

dazu:

<https://docs.google.com/document/d/1wCSor0EFP3Uk3nQKOeKKBkoXrLFXRKQoaRIAtBPqPQg/edit?usp=sharing>

B, welche wiederum Teil der Sammlungen C und D ist), es kann beispielsweise auch der institutionelle und administrative Kontext der Sammlung dokumentiert werden. Aufgrund der starken Einbeziehung bereits etablierter Terminologie aus dem Dublin Core Universum kann außerdem davon ausgegangen werden, dass ein Großteil der zur Verwendung kommenden Elemente bekannt ist bzw. deren Bedeutung auch ohne allzu intensive Einarbeitung verstanden werden kann. Sollte dies im Einzelfall nicht immer möglich sein, so kann immer noch auf die jeweils umfassenden Dokumentationen zurückgegriffen werden.

Wie vor allem der Use-Case zu DCAP zeigt, handelt es sich bei DCCAP um eine solide Ausgangsbasis für die Erstellung eines Datenmodells zur Sammlungsbeschreibung. Wie das Beispiel DCAP jedoch auch in aller Deutlichkeit zeigt, steckt DCCAP nur den Rahmen des Datenmodells fest und stellt eine breite Palette nachnutzbarer Elemente zur Verfügung, ist aber selbst kein Datenmodell.

3. Fazit

Klammert man METS und PREMIS einmal aus, so kann festgehalten werden, dass mit all den übrigen in Kapitel 2 besprochenen Standards Sammlungen durchaus detailliert und umfassend beschreibbar sind. Dies dürfte insofern kaum überraschen, da diese Standards dahingehend konzipiert wurden, Objekte aus dem Kontext von Gedächtnisorganisationen zu beschreiben bzw. um deskriptive/identifizierende wie administrative Metadaten solcher Objekte zu dokumentieren.⁵⁰ So unterschiedlich die einzelnen Objekte dabei im Einzelnen auch sein mögen, ist die Schnittmenge ihrer überlieferungswürdigen Eigenschaften doch beträchtlich. So hat ein Objekt meist einen Namen, eine Form, eine Größe und ein Entstehungsdatum. Es kann außerdem häufig mit einer Person, einem Erzeuger oder Schöpfer in Verbindung gebracht werden, ebenso wie mit einem Besitzer und/oder Verwalter bzw. einer entsprechenden Einrichtung oder Organisation. Neben der zeitlichen Verortung ist in vielen Fällen auch eine räumliche gegeben, wie auch eine thematische Verschlagwortung.

Dieser keinesfalls vollständigen Auflistung weitgehend objektunabhängiger oder genereller Eigenschaften stehen natürlich objektspezifische Eigenschaften und Besonderheiten gegenüber. Hinzu kommen noch verschiedene Anwendungsszenarien, welche differenzierte Anforderungen an einen Metadatenstandard zur Beschreibung von Objekten aus dem weiten Feld der Gedächtnisorganisationen stellen.

Während TEI, EAD, LIDO oder DCCAP vornehmlich deskriptive und administrative Metadaten abdecken, liegt der Schwerpunkt von METS und PREMIS in erster Linie auf technischen Metadaten, die vor allem für das Zusammenspiel mit Informationsverwaltungssystemen benötigt werden. Informationssystemen, die zur Anzeige, Suche, (Langzeit-)Archivierung oder für Austausch von digitalen Objekten,

⁵⁰ Stefanie Rühle, *Kleines Handbuch Metadaten. Metadaten* (Göttingen: KIM - Kompetenzzentrum Interoperable Metadaten, 2012), 2, <http://www.kim-forum.org/Subsites/kim/SharedDocs/Downloads/DE/Handbuch/metadatenprofile.html>.

einerlei ob es sich dabei um digitale Derivate physischer Objekte oder um „born digital“-Ressourcen handelt, konzipiert wurden.

Die Frage nach dem 'richtigen' Metadatenstandard für die Beschreibung von Objekten hängt also sowohl vom Typ des Objektes selbst ab, als auch von dem konkreten Anwendungsfall, für den man eine standardisierte und maschinenlesbare Sammlungsbeschreibung benötigt. Ist man vornehmlich an der Dokumentation deskriptiver wie administrativer Metadaten interessiert, so bietet sich für eine Beschreibung eines Objektes vom Typ Sammlung in erster Linie wohl eine auf dem DCCAP basierendes Datenmodell an, da DCCAP ja konkret für die Beschreibung von Sammlungen entworfen wurde. Dabei gilt es aber zu bedenken, dass DCCAP bewusst für 'einfache' Beschreibungen für Sammlungen unterschiedlichster Natur entworfen wurde, worauf die Verfasser des Dublin Core Collection Application Profiles auch dezidiert hinweisen: „The DC Collection AP is intended to provide a means of creating **simple** descriptions [...] suitable for a broad range of collection“.⁵¹

3.1. Ausblick

Angesichts der Zielsetzung von DARIAH-DE Cluster-4 „Wissenschaftliche Sammlungen“, die Sichtbarkeit solcher wissenschaftlichen Sammlungen durch die Schaffung eines entsprechenden zentralen Nachweissystems zu erhöhen, ist ein auf DCCAP basierender Metadatenstandard die wohl beste Wahl zur Beschreibung solcher wissenschaftlicher Sammlungen. Die Modellstruktur ist einerseits wenig komplex, andererseits aber flexibel genug, die unterschiedlichsten Sammlungen in vergleichbarer Art und Weise, maschinen- als auch menschenlesbar zu beschreiben.

Die für solche Beschreibungen notwendigen Schritte sind darüber hinaus schon in der ersten Projektphase von DARIAH-DE umgesetzt worden. Konkret handelt es sich dabei um die weit fortgeschrittene Spezifizierung eines auf DCCAP basierenden Datenmodells. Dieses Datenmodell gilt es nun weiter auszubauen und noch näher zu spezifizieren. Ist dieser Schritt abgeschlossen, sollte eine Serialisierung des Modells keine allzu große intellektuelle Herausforderung mehr darstellen. Sobald eine solche Serialisierung als XML-Schema vorliegt, können Crosswalks definiert werden, mit deren Hilfe Sammlungsbeschreibungen, die in den übrigen in Kapitel 2 beschriebenen Metadatenstandards vorliegen, auf das DARIAH Modell von DCCAP überführt werden können, und umgekehrt.

Der aktuelle Entwicklungsstand des **DARIAH Collection Description Data Models** (DCDDM) samt Serialisierung nach XML-Schema kann auf GitHub verfolgt werden.

<https://github.com/csae8092/DCDDM>

Kritik, Kommentare, Anregungen sind nachdrücklich erwünscht.

⁵¹ *Dublin Core Collections Application Profile* (Dublin Core Collection Description Task Group, March 9, 2007), <http://dublincore.org/groups/collections/collection-application-profile/>.

4. Literaturverzeichnis

- ATHENA Project. *Lightweight Information Describing Objects (LIDO): The International Harvesting Standard for Museums*. Rom, 2011. [http://www.athe-
naeurope.org/getFile.php?id=786](http://www.athe-
naeurope.org/getFile.php?id=786).
- Caplan, Priscilla. *PREMIS Verstehen*. Translated by Tobias Beinert, 2009. http://www.loc.gov/standards/premis/understanding_premis_german.pdf.
- Coburn, Erin, Richard Light, Gordon McKenna, Regine Stein, and Axel Vitzthum. *LIDO - Lightweight Information Describing Objects Version 1.0*, 2010.
- Corsten, Severin, and Claus W. Gerhardt, eds. *Lexikon Des Gesamten Buchwesens. Bd. 6: Phraseologie - Schütz. 2.*, völlig Neubearb. Aufl. Stuttgart: Hiersemann, 2003.
- Crofts, Nick, Martin Doerr, Tony Gill, Stephen Stead, and Matthew Stiff. *Definition of the CIDOC Conceptual Reference Model*, 2009. <http://www.cidoc-crm.org/crm-concepts/>.
- DFG - Deutsche Forschungsgemeinschaft. "DFG-Praxisregeln 'Digitalisierung,'" February 2013. http://www.dfg.de/formulare/12_151/12_151_de.pdf.
- Digital Library Federation. *<METS> Metadata Encoding and Transmission Standard: Primer and Reference Manual*, 2010. <http://www.loc.gov/standards/mets/METSPrimerRevised.pdf>.
- Dublin Core Collections Application Profile*. Dublin Core Collection Description Task Group, March 9, 2007. <http://dublincore.org/groups/collections/collection-application-profile/>.
- Dublin Core Collections Application Profile*. Dublin Core Collection Description Task Group, March 9, 2007. <http://dublincore.org/groups/collections/collection-application-profile/>.
- Glück, Helmut, ed. *Metzler-Lexikon Sprache. 2.*, überarb. und erw. Aufl. Stuttgart [u.a.]: Metzler, 2000.
- Löbnitz, Anke, and Angelika Menne-Haritz. *Encoded Archival Description Tag-Library. Version 2002, Übersetzt von Anke Löbnitz Und Angelika Menne-Haritz*. Berlin, 2006. <http://www.bundesarchiv.de/imperia/md/content/daofind/1.pdf>.
- Menne-Haritz, Angelika, and Michael J. Fox, trans. *Das EAD-Kochbuch. (ohne Abschnitte 4 Und 5) Übersetzung Des EAD Cook Book (Vers. 1) von Michael J. Fox*, 2003. <http://www.bundesarchiv.de/imperia/md/content/daofind/eadkochbuch.pdf>.
- Meyer, Sebastian, Carolin Schreiber, Thomas Busch, Robert Giel, Christoph Mackert, and Torsten Schaßan. *TEI-Anwendungsprofil Für Digitalisierte Handschriften*, 2014.
- PREMIS Editorial Committee. *Introduction and Supporting Materials from PREMIS Data Dictionary for Preservation Metadata*, 2012. <http://www.loc.gov/standards/premis/v2/premis-report-2-2.pdf>.
- . *PREMIS Data Dictionary for Preservation Metadata*, 2012.
- RLG. *RLG Best Practice Guidelines for Encoded Archival Description*. Mountain View, California, 2002. <http://www.oclc.org/content/dam/research/activities/ead/bpg.pdf?urlm=161431>.
- Rühle, Stefanie. *Kleines Handbuch Metadaten. Metadaten*. Göttingen: KIM - Kompetenzzentrum Interoperable Metadaten, 2012. <http://www.kim-forum.org/Subsites/kim/SharedDocs/Downloads/DE/Handbuch/metadatenprofile.html>.

The Consultative Committee for Space Data Systems (CCSDS). *Reference Model for an Open Archival Information System (OAIS)*. Recommended Practices. Washington D.C., 2012. <http://public.ccsds.org/publications/archive/650x0m2.pdf>.
“What Is LIDO.” Accessed February 22, 2013. <http://network.icom.museum/cidoc/working-groups/data-harvesting-and-interchange/what-is-lido/>.

5. Abbildungsverzeichnis

Abbildung 1: Entstehungsdatum eines Objektes in LIDO. 16