

H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, M. Jehn (Hrsg.)

nestor Handbuch

Eine kleine Enzyklopädie
der digitalen Langzeitarchivierung

Version 2.0

Kapitel 6
Metadatenstandards
im Bereich der digitalen LZA

nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung
hg. v. H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, M. Jehn
im Rahmen des Projektes: nestor – Kompetenznetzwerk Langzeitarchivierung und
Langzeitverfügbarkeit digitaler Ressourcen für Deutschland
nestor – Network of Expertise in Long-Term Storage of Digital Resources
<http://www.langzeitarchivierung.de/>

Kontakt: editors@langzeitarchivierung.de
c/o Niedersächsische Staats- und Universitätsbibliothek Göttingen,
Dr. Heike Neuroth, Forschung und Entwicklung, Papendiek 14, 37073 Göttingen

Die Herausgeber danken Anke Herr (Korrektur), Martina Kerzel (Bildbearbeitung) und
Jörn Tietgen (Layout und Formatierung des Gesamttextes) für ihre unverzichtbare
Unterstützung bei der Fertigstellung des Handbuchs.

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen
Nationalbibliografie; detaillierte bibliografische Daten sind im Internet unter
<http://www.d-nb.de/> abrufbar.

Die Inhalte dieses Buchs stehen auch als Onlineversion
(<http://nestor.sub.uni-goettingen.de/handbuch/>)
sowie über den Göttinger Universitätskatalog (<http://www.sub.uni-goettingen.de>) zur
Verfügung.

Die digitale Version 2.0 steht unter folgender Creative-Commons-Lizenz:
„Attribution-Noncommercial-Share Alike 3.0 Unported“
<http://creativecommons.org/licenses/by-nc-sa/3.0/>



Einfache Nutzungsrechte liegen beim Verlag Werner Hülsbusch, Boizenburg.
© Verlag Werner Hülsbusch, Boizenburg, 2009
www.vwh-verlag.de
In Kooperation mit dem Universitätsverlag Göttingen

Markenerklärung: Die in diesem Werk wiedergegebenen Gebrauchsnamen, Handelsnamen,
Warenzeichen usw. können auch ohne besondere Kennzeichnung geschützte Marken sein und
als solche den gesetzlichen Bestimmungen unterliegen.

Druck und Bindung: Kunsthaus Schwanheide

Printed in Germany – Als Typoskript gedruckt –

ISBN: 978-3-940317-48-3

URL für Kapitel 6 „Metadatenstandards im Bereich der digitalen LZA“ (Version 2.0):
<urn:nbn:de:0008-20090811252>
<http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0008-20090811252>



Gewidmet der Erinnerung an Hans Liegmann (†), der als Mitinitiator und früherer Herausgeber des Handbuchs ganz wesentlich an dessen Entstehung beteiligt war.

6 Metadatenstandards im Bereich der digitalen LZA

6.1 Einführung

Matthias Jehn

Für den Erfolg der digitalen Langzeitarchivierung bilden Standards eine unabdingbare Voraussetzung für kompatible und interoperative Systeme aller Art. Sie werden für technische als auch für organisatorische Aspekte in der digitalen Langzeitarchivierung benötigt. Ein Standard kann in einem formalisierten oder nicht-formalisierten Regelwerk bzw. in einem sich ungeplant ergebenden Regelfall bestehen, beispielsweise in einer einzelnen Regel bzw. mehreren Regeln oder einer Norm. Standards fördern nicht nur die Wiederverwendbarkeit und Austauschbarkeit von Komponenten, sondern gewähren auch verlässliche Vorgaben für System- und Produktentwickler. Öffentlich verfügbare und realistisch umsetzbare Vorgaben sind Basis für konkurrierende Implementierungen und somit für einen funktionierenden Markt. Das notwendige Konsensprinzip erfordert dabei sehr aufwändige Abstimmungsprozesse und wirtschaftliche oder

sonstige Interessen führen teilweise zu konkurrierenden Inhalten oder unnötigem Umfang von Ansätzen. Die Abgrenzung von Inhalten und die zeitliche Synchronisation können zudem auch durch die Vielzahl der Standardisierungsorganisationen negativ beeinflusst werden. Auf jeden Fall ist das Prozedere der Standardisierung und der Aufbau der Standards sehr unterschiedlich. Die geforderte Offenheit von Standards ist nicht nur eine rein definitorische Angelegenheit, sondern kann weitgehende rechtliche und wirtschaftliche Konsequenzen haben. Versteckte Patente oder sonstige Hindernisse, die z.B. Mitbewerber bei einer Implementierung behindern, können sich nachteilig auf die Zuverlässigkeit und Wirtschaftlichkeit der Langzeitarchivierung auswirken. Vorteilhaft ist, dass sich die Standardisierungsorganisationen um mehr Transparenz und auch Einheitlichkeit bei der Behandlung und Darstellung von Rechten (Intellectual Property Rights – IPR) bemühen. Das folgende Kapitel präsentiert einige wesentliche Entwicklungen im Bereich der internationalen Standards und der Bemühungen, im Bereich der technischen Standards und der Metadatenstandards für die digitale Langzeitarchivierung zu entwickeln.

6.2 Metadata Encoding and Transmission Standard – Einführung und Nutzungsmöglichkeiten

Markus Enders

Ausgehend von den Digitalisierungsaktivitäten der Bibliotheken Mitte der 1990er Jahre entstand die Notwendigkeit, die so entstandenen Dokumente umfassend zu beschreiben. Diese Beschreibung muss im Gegensatz zu den bis dahin üblichen Verfahrensweisen nicht nur einen Datensatz für das gesamte Dokument beinhalten, sondern außerdem einzelne Dokumentbestandteile und ihre Abhängigkeiten zueinander beschreiben. So lassen sich gewohnte Nutzungsmöglichkeiten eines Buches in die digitale Welt übertragen. Inhaltsverzeichnisse, Seitennummern sowie Verweise auf einzelne Bilder müssen durch ein solches Format zusammengehalten werden.

Zu diesem Zweck wurde im Rahmen des „Making Of Amerika“ Projektes Ebind entwickelt¹. Ebind selber war jedoch ausschließlich nur für Digitalisate von Büchern sinnvoll zu verwenden.

Um weitere Medientypen sowie unterschiedliche Metadatenformate einbinden zu können, haben sich Anforderungen an ein komplexes Objektformat ergeben. Dies setzt ein abstraktes Modell voraus, mit Hilfe dessen sich Dokumente flexibel modellieren lassen und als Container Format verschiedene Standards eingebunden werden können. Ein solches abstraktes Modell bildet die Basis von METS und wird durch das METS-XML-Schema beschrieben. Daher wird METS derzeit auch fast ausschließlich als XML serialisiert und in Form von Dateien gespeichert. Als Container Format ist es in der Lage weitere XML-Schema (so genannte Extension Schemas) zu integrieren.

Das METS Abstract Model

Das METS „Abstract Model“ beinhaltet alle Objekte innerhalb eines METS Dokuments und beschreibt deren Verhältnis zueinander. Zentraler Bestandteil eines METS-Dokuments ist eine Struktur. Das entsprechende Element nennt sich daher structMap und ist als einziges Element im „Abstract Model“ verpflichtend. Jedes METS Dokument muss ein solches Element besitzen. Unter Struktur wird in diesem Fall eine hierarchische Struktur mit nur einem Start-

1 O.V.: An Introduction to the Electronic Binding DTD (Ebind). <http://sunsite.berkeley.edu/Ebind/>

Alle hier aufgeführten URLs wurden im April 2009 auf Erreichbarkeit geprüft .

knoten verstanden. Eine Struktur kann also als Baum interpretiert werden. Der Wurzelknoten sowie jeder Ast wird als Struktureinheit bezeichnet. Jede Struktur muss über einen Wurzelknoten verfügen. In der Praxis kann diese verpflichtende Struktur bspw. die logische Struktur – also das Inhaltsverzeichnis einer Monographie speichern. Im Minimalfall wird dieses lediglich die Struktureinheit der Monographie umfassen, da der Wurzelknoten in dem Baum verpflichtend ist. Weitere Strukturen sind optional. Eine weitere Struktur könnte bspw. die physische Struktur des Dokuments sein. Die physische Struktur beschreibt bspw. aus der Exemplarsicht (gebundene Einheit mit Seiten als unterliegende Struktureinheiten).

Verknüpfungen zwischen zwei Struktureinheiten werden in einer separaten Sektion gespeichert. Das „Abstract Model“ stellt dazu die structLink Sektion zur Verfügung, die optional genutzt werden kann. Jede Verknüpfung zwischen zwei Struktureinheiten wird in einem eigenen Element definiert.

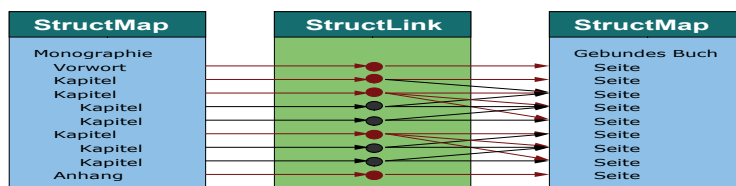


Abbildung 1: Verknüpfung von zwei Strukturen im Abstract-Model

Das „Abstract Model“ macht allerdings keine Vorgaben, aus welcher strukturellen Perspektive ein Dokument beschrieben wird oder wie detailliert diese einzelnen Strukturen ausgearbeitet werden müssen.

Ferner berücksichtigt das „Abstract Model“ auch Metadaten. Hierunter sind allerdings nicht nur bibliographische Metadaten zu verstehen. Vielmehr wird in deskriptive Metadaten (in der Descriptive Metadata Section) und administrative Metadaten (in der Administrative Metadata Section) unterschieden. Während die deskriptiven Metadaten bibliographische Informationen enthalten, werden Informationen zu Rechteinhabern, Nutzungsrechte, technische Informationen zu einzelnen Dateien oder Langzeitarchivierungsmetadaten in den administrativen Metadaten gespeichert. Für beide Metadatentypen können beliebige Schema, so genannte „Extension Schema“ genutzt werden, die in der jeweiligen Sektion gespeichert werden. Auch die Referenzierung von Metadatensätzen ist möglich, sofern diese bspw. per URL zugänglich sind. Jede Datei sowie jeder Struktureinheit lässt sich mit entsprechenden Metadatensätzen versehen, wobei jeder Einheit mehrere Datensätze zugeordnet werden können. Als „Extensi-

on Schema“ können sowohl XML-Metadatenschema wie bspw. MARC XML, MODS, Dublin Core) sowie Binärdaten benutzt werden. Dies erlaubt auch die Integration gängiger bibliothekarischer Standards wie bspw. PICA-Datensätze.



Abbildung 2: Verweis auf Metadatensektionen im METS-Abstract-Model

Neben den Struktureinheiten und ihren zugehörigen Metadaten spielen auch Dateien bzw. Streams eine wesentliche Rolle, da letztlich in ihnen die durch das METS-Dokument beschriebenen Inhalte manifestiert/gespeichert sind. Eine Datei kann bspw. den Volltext eines Buches, die Audioaufnahme einer Rede oder eine gescannte Buchseite als Image enthalten. Entsprechende Daten können in ein METS-Dokument eingebunden werden (bspw. Base64 encoded in die METS-XML Datei eingefügt werden) oder aber mittels xlink referenziert werden. Ein METS-Dokument kann also als Container alle für ein Dokument notwendigen Dateien enthalten oder referenzieren, unabhängig davon, ob die Dateien lokal oder auf entfernten Servern vorhanden sind. Metadatensätze, die nicht in die METS Datei eingebunden sind, werden nicht als Datei betrachtet, sondern sind aus der entsprechenden Metadatensektion zu referenzieren.

Grundsätzlich müssen alle für ein METS-Dokument relevanten Dateien innerhalb der File-Sektion aufgeführt werden. Innerhalb der File-Sektion können Gruppen (File-Groups) von Dateien gebildet werden, wobei die Abgrenzungskriterien zwischen einzelnen Gruppen nicht durch das „Abstract Model“ definiert sind. Je nach Modellierung lassen sich Dateien bspw. nach technischen Parametern (Auflösung oder Farbtiefe von Images), Anwendungszweck (Anzeige, Archivierung, Suche) oder sonstigen Eigenschaften (Durchlauf bestimmter Produktionsschritte) den einzelnen Gruppen zuordnen.

Das METS-Abstract-Model erlaubt das Speichern von administrativen Metadaten zu jeder Datei. Generelle, für jede Datei verfügbare technische Metadaten wie Dateigröße, Checksummen etc. lassen sich direkt in METS speichern. Für weiterführende Metadaten kann mit jeder Datei eine oder mehrere Administrative Metadatensektion(en) verknüpft werden, die bspw. Formatspezifische Metadaten enthalten (für Images könnten die Auflösungsinformationen, Informationen zur Farbtiefe etc. sein).

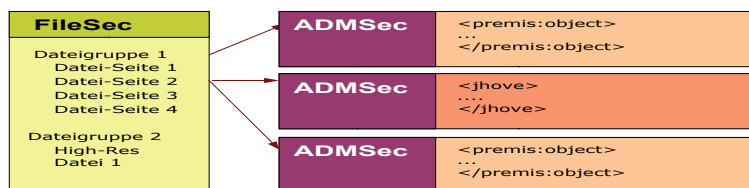


Abbildung 3: Administrative Metadata zu Dateien

Dateien sind darüber hinaus mit Struktureinheiten verknüpft. Die Struktureinheit, die eine einzelne Buchseite repräsentiert, kann somit mit einer einzelnen Datei, die ein Image dieser Seite beinhaltet, verknüpft werden. Das „METS-Abstract-Model“ stellt hierzu eine N:M Verknüpfung bereit. Das bedeutet, dass eine Datei von mehreren Struktureinheiten (auch aus unterschiedlichen Struktursektionen) aus verknüpft werden kann, genauso wie eine Struktureinheit mehrere Dateien verknüpfen kann. Im Ergebnis heißt das, dass der Struktureinheit vom Typ „Monographic“ sämtliche Imagedateien eines gesamteten Werkes direkt unterstellt sind.

Für die Verknüpfung von Dateien sieht das „METS-Abstract-Model“ noch weitere Möglichkeiten vor. So lassen sich mehrere Verknüpfungen hinsichtlich ihrer Reihenfolge beim Abspielen bzw. Anzeigen bewerten. Dateien können entweder sequentiell angezeigt (Images eines digitalisierten Buches) oder auch parallel abgespielt (Audio- und Videodateien gleichen Inhalts) werden. Darüber hinaus kann nicht nur auf Dateien, sondern auch in Dateiobjekte hinein verlinkt werden. Diese Verlinkungen sind u.a. dann sinnvoll, wenn Einheiten beschrieben werden, die aus technischen Gründen nicht aus der Datei herausgetrennt werden können. Das können bestimmte Teile eines Images sein (bspw. einzelne Textspalten) oder aber konkrete zeitliche Abschnitte einer Audioaufnahme. In der Praxis lassen sich so einzelne Zeitabschnitte eines Streams markieren und bspw. mit inhaltlich identischen Abschnitten eines Rede-Manuskriptes taggen. Das METS-Dokument würde über die Struktureinheit eine Verbindung zwischen den unterschiedlichen Dateien herstellen.

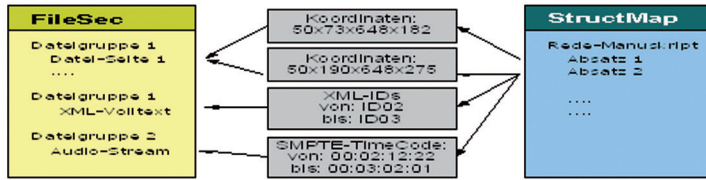


Abbildung 4: Struktureinheit ist mit verschiedenen Dateien und Dateibereichen verknüpft

Das METS-Abstract-Model nutzt intensiv die Möglichkeit, einzelne Sektionen miteinander zu verknüpfen. Da METS überwiegend als XML realisiert ist, geschieht diese Verknüpfung über XML-Identifizier. Jede Sektion verfügt über einen Identifizier, der innerhalb des XML- Dokumentes eindeutig ist. Er dient als Ziel für die Verknüpfungen aus anderen Sektionen heraus. Aufgrund der XML-Serialisierung muß er den XML-ID Anforderungen genügen. Es muss bei Verwendung von weiteren Extension Schemas darauf geachtet werden, dass die Eindeutigkeit der Identifizier aus dem unterschiedlichen Schema nicht gefährdet wird, da diese üblicherweise alle im gleichen Namensraum existieren.

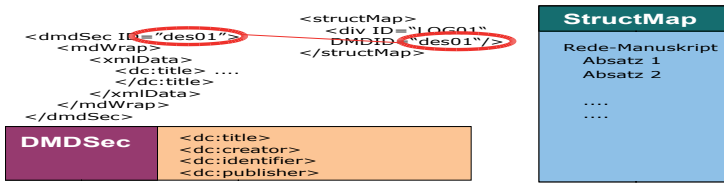


Abbildung 5: Unterschiedliche Sektionen mittels XML-IDs verknüpft

Dokumentation

Wie deutlich geworden ist, stellt das METS-Abstract-Model sowie des XML-Serialisierung als METS-XML Schema lediglich ein grobes Modell da, welches auf den jeweiligen Anwendungsfall angepasst werden muss. Die Verwendung von Extension Schema sollte genauso dokumentiert werden wie die Nutzung optionaler Elemente und Attribute in METS. Hierbei sollte vor allem auch die Transformation realer, im zu beschreibenden Dokument vorhandene Objekte in entsprechende METS-Objekte bzw. METS-Sektionen im Vordergrund stehen. Eine einzige Strukturektion kann bspw. logische Einheiten (bspw. das Inhaltsverzeichnis eines Buches) umfassen als auch bestimmte physische Einheiten (bspw. einzelne Seiten) enthalten. Alternativ können jedoch bestimmte Einheiten in eine separate Strukturektion ausgelagert werden. Das „Abstract

Model“ erlaubt diese Flexibilität. Eine Implementierung von METS für einen bestimmten Anwendungsfall muss dieses jedoch konkret festlegen.

Um die Dokumentation zu standardisieren wurde das METS-Profil Schema entwickelt. Es gibt eine Grobstrukturierung vor, die sicher stellt, dass alle wesentlichen Bereiche eines METS-Dokuments in der Dokumentation berücksichtigt werden. Die Dokumentation selber muss derzeit noch auf XML Basis erfolgen. Die so entstandene XML-Datei lässt sich jedoch anschliessend als HTML oder PDF konvertieren.

Um ein solches Profil auf der offiziellen METS-Homepage veröffentlichen zu können, wird es durch Mitglieder des METS-Editorial-Board verifiziert. Nur verifizierte METS-Profile werden veröffentlicht und stehen auf der Homepage zur Nachnutzung bereit. Sie können von anderen Institutionen adaptiert und modifiziert werden und somit erheblich zur Reduktion der Entwicklungszeit einer eigenen METS-Implementierung beitragen.

Fazit

Aufgrund der hohen Flexibilität des METS Abstract Models wird METS in einer großen Zahl unterschiedlicher Implementierungen für sehr verschiedene Dokumententypen genutzt. Neben der ursprünglichen Anwendung, digitalisierte Büchern zu beschreiben, existieren heute sowohl METS-Profile zur Webseitenbeschreibungen (Webarchivierung) sowie Audio- und Videodaten. Während in den ersten Jahren METS überwiegend zum Beschreiben komplexer Dokumente genutzt wurde, um diese dann mittels XSLTs oder DMS-Systeme verwalten und anzeigen zu können, wird METS heute gerade auch im Bereich der Langzeitarchivierung zur Beschreibung des Archival Information Packets (AIP) genutzt. METS ist heute für viele Bereiche, in denen komplexe Dokumente beschrieben werden müssen, ein De-facto-Standard und kann sowohl im universitären als auch im kommerziellen Umfeld eine große Zahl an Implementierungen vorweisen. Ein großer Teil derer ist im METS-Implementation Registry auf der METS-Homepage (<http://www.loc.gov/mets>) nachgewiesen.

6.3 PREMIS

Olaf Brandt

Das Akronym PREMIS löst sich in „PREservation Metadata: Implementation Strategies“ auf. PREMIS ist eine Initiative, welche die Entwicklung und Pflege des international anerkannten gleichnamigen PREMIS-Langzeitarchivierungsmetadatenstandards verantwortet. Sie wurde im Jahre 2003 von OCLC (Online Computer Library Center) und RLG (Research Library Group) ins Leben gerufen.

Langzeitarchivierungsmetadaten sind - vereinfacht ausgedrückt - strukturierte Informationen über digitale Objekte, ihre Kontexte, ihre Beziehungen und Verknüpfungen, welche die Prozesse der digitalen Langzeitarchivierung ermöglichen, unterstützen oder dokumentieren.

Das Hauptziel von PREMIS ist die Entwicklung von Empfehlungen, Vorschlägen und Best-Practices zur Implementierung von Langzeitarchivierungsmetadaten, d.h. die Fortentwicklung des Standards, sowie die Anbindung an weitere Standards. Die Fortentwicklung wird zurzeit vom PREMIS Editorial Committee geleistet. Das ist eine internationale Gruppe von Akteuren aus Gedächtnisorganisationen wie Archiven, Bibliotheken und Museen sowie der Privatwirtschaft. Die Arbeit von PREMIS baut auf den Ergebnissen der Preservation-Metadata Working-Group auf, die bereits 2001 die Entwicklung eines gemeinsamen Rahmenkonzeptes für Langzeitarchivierungsmetadaten vorantrieb.² Nach der Veröffentlichung des PREMIS Data Dictionaries der Version 1.0 im Jahr 2005³ galt es zunächst Implementierungen zu unterstützen und die Weiterentwicklung von PREMIS zu institutionalisieren. Dafür wurde eine PREMIS Managing Agency gegründet, welche an der Library of Congress angesiedelt ist.⁴ Sie übernimmt in enger Abstimmung mit dem PREMIS Editorial Committee die Koordination von PREMIS im Hintergrund. Zu den Aufgaben gehören z.B. das Hosting und die Pflege der Webseite, die Planung und Durchführung von Maßnahmen für die PREMIS-Verbreitung und der Betrieb und die Moderation der PREMIS-Diskussionslisten. Das PREMIS Editorial Committee erarbeitet zusammen mit der Managing

2 Preservation Metadata Working Group (PMWG 2002) Framework:

http://www.oclc.org/research/projects/pmwg/pm_framework.pdf

3 Abschlußbericht der PREMIS Arbeitsgruppe mit „Data Dictionary for Preservation Metadata“: <http://www.oclc.org/research/projects/pmwg/premis-final.pdf>

4 Webseite der PREMIS Maintenance Activity: <http://www.loc.gov/standards/premis/>

Agency die Ziele und die weitere Entwicklung von PREMIS. Das betrifft v.a. die Weiterentwicklung und Pflege des Data Dictionary und der XML-Schemas. Weiter sorgt das Editorial Committee für die Verbreitung des Wissens über PREMIS durch Vorträge und Publikationen. Die PREMIS Implementors Group ist über eine Mailingliste und ein Wiki organisiert. Sie ist offen für jede Person oder Institution, die ein Interesse an digitaler Langzeitarchivierung oder PREMIS hat.

Wichtigste Neuerung des Jahres 2008 ist sicherlich die Veröffentlichung des PREMIS Data Dictionary für Langzeitarchivierungsmetadaten in Version 2.0 und des neu erarbeiteten generischen XML-Schemas.⁵ Aber auch die Fortschritte bei der Implementierung von PREMIS und METS sind in ihrer Bedeutung sicherlich nicht zu unterschätzen. So ist PREMIS seit einiger Zeit ein offizielles Erweiterungsschema von METS.⁶ Empfehlungen für die Implementierung von PREMIS und METS⁷ finden ihren Niederschlag in fruchtbaren Diskussionen.⁸ PREMIS hat sich in der Langzeitarchivierungscommunity einen festen Platz als Nachschlagewerk für Implementierungen von Langzeitarchivierungsmetadaten und als gemeinsames Austauschformat⁹ erarbeitet.

Um einen ersten Einblick in die Welt von PREMIS zu bekommen, wird im nun folgenden Abschnitt eine Einführung in das PREMIS-Datenmodell gegeben.

Aufbau Datenmodell

Das PREMIS Datenmodell kennt fünf grundlegende Einheiten, sog. Entities:

- Intellectual Entities
- Object Entity
- Events Entity
- Rights Entity
- Agent Entity

5 Siehe dazu <http://www.loc.gov/standards/premis/v2/premis-2-0.pdf> und <http://www.loc.gov/standards/premis/schemas.html>

6 Siehe dazu <http://www.loc.gov/standards/mets/mets-extenders.html>

7 Siehe dazu <http://www.loc.gov/premis/guidelines-premismets.pdf>

8 Siehe dazu <http://www.dlib.org/dlib/september08/dappert/09dappert.html>

9 Siehe dazu http://www.library.cornell.edu/dlit/MathArc/web/resources/MathArc_metadataschema031a.doc oder auch in jüngster Zeit <http://www.dlib.org/dlib/november08/caplan/11caplan.html>

Entities sind abstrakte Klassen von 'Dingen', also z.B. „digitale Objekte“ oder „Agenten“. Die Eigenschaften von vier Entities werden im PREMIS Data Dictionary mit sog. Semantic Units (semantische Einheiten) näher beschrieben. Semantic Units sind die für die digitale Langzeitarchivierung relevanten Eigenschaften der Entities.

Intellectual Entities

Intellectual Entities sind als zusammenhängende Inhalte definiert, die als eine Einheit beschrieben werden. Sie stellen somit eine Idee dar, welche in analogen oder digitalen Manifestationen oder Repräsentationen vorliegen kann. Es könnte sich also sowohl um einen Zeitschriftenband handeln als auch um den digitalisierten Zeitschriftenband. Dieser kann wiederum weitere Intellectual Entities (z.B. Zeitschriftenausgaben oder Artikel) enthalten. Intellectual Entities werden im Data Dictionary nicht mit semantischen Einheiten beschrieben, da sie außerhalb des Fokus, Kerninformationen für die digitale Langzeitarchivierung bereitzustellen, liegen. Auf sie kann aber von Objekten verwiesen werden.

Object Entity

In der *Object Entity* werden die zu archivierenden Daten mit relevanten Informationen für das Management und die Planung von Langzeitarchivierungsprozessen beschrieben. Die Object Entity kann unterschiedliche digitale Objekte beschreiben: sogenannte Representations, Dateien und auch Bitstreams.

Eine *Representation* ist eine Instanz oder Manifestierung einer Intellektuellen Entität, realisiert oder enthält sie also. Eine Representation ist eine logisch-funktionale Einheit aus digitalen Daten oder Dateien und ihrer Strukturbeschreibung. Als Beispiel kann eine Webseite dienen, die aus mehreren einzelnen Dateien besteht. Ihre Struktur und die Beziehungen der einzelnen Elemente untereinander zu kennen ist essentiell für die langfristige, sinnvolle und komplette Darstellung dieser Webseite als Einheit. Beim gegebenen Beispiel einer Webseite müsste z.B. beschrieben werden, dass eine Einstiegsseite existiert, die auf bestimmte Art und Weise mehrere Unterseiten und andere Elemente (wie z.B. Grafikdateien) einbindet. Dateien werden im PREMIS Data Dictionary als „named and ordered sequence of bytes that is known by an operating system“ bezeichnet. *Bitstream* (Datenstrom) wird nur als in zu archivierenden Dateien enthaltener und adressierbarer Teil beschrieben. Ein Datenstrom kann nur durch Umwandlung oder Hinzufügung von weiteren Informationen zu einer Datei werden. Zu den beschreibbaren Informationen von Objekten gehören z.B. eindeutige Identifikatoren, Charakteristika der Daten wie Größe und Format, Beschrei-

bungen der Systemumgebungen (Software, Hardware), Beschreibungsmöglichkeiten der relevanten Eigenschaften der Objekte, sowie die Beziehungen zu anderen Objekten, Events und Rechteinformationen.

Event Entity

Ein *Event* ist in PREMIS eine identifizierbare Aktion oder ein Ereignis, in das mindestens ein Objekt und/oder ein Agent einbezogen sind. In der *Event Entity* werden Informationen über diese Aktionen oder Ereignisse und ihre Resultate sowie ihre Beziehungen zu Objekten und Agenten beschrieben. Mit der lückenlosen Aufzeichnung der Ereignisse im Archiv kann die Geschichte und die Verwendung der digitalen Objekte im Archivsystem nachgewiesen werden. Die Dokumentation der Ereignisse dient also dem Nachweis der Provenienz, als Beleg für die Einhaltung von Rechten oder kann für Statistikfunktionen und Billing herangezogen werden.

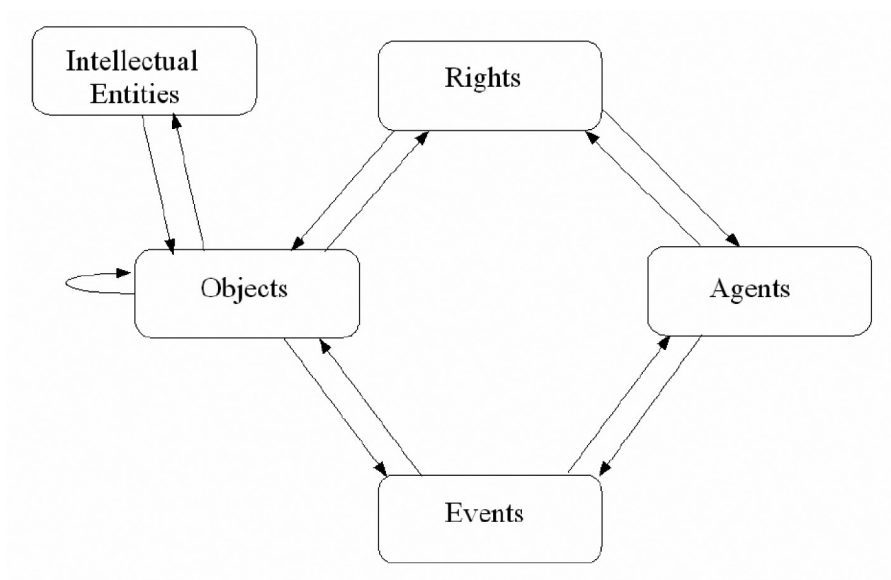
Agent Entity

Ein PREMIS *Agent* ist definiert als Person, Organisation oder Software, welche auf ein Ereignis im digitalen Archiv bezogen ist. Mit der *Agent Entity* werden spezifische Informationen von Agenten beschrieben, die im Zusammenhang mit Langzeitarchivierungsereignissen und Rechtemanagement im Leben eines Datenobjektes auftreten. Informationen über Agenten dienen v.a. der eindeutigen Identifizierung eines Agents.

Rights-Entity

Für den Betrieb eines Langzeitarchivs ist es wichtig, von den mit den Objekten verbundenen Rechten mit Relevanz für die Planung und Durchführung von Aktionen für die digitale Langzeitarchivierung zu wissen. Das betrifft z.B. das Kopieren von Daten, die Umwandlung in andere Formate etc. Aussagen über diese Rechte und Erlaubnisse werden in der *Rights Entity* beschrieben. Seit PREMIS 2.0 können tiefer gehende Rechtekonstellationen und deren Kontexte beschrieben werden, wie z.B. spezifische Urheberrechte in einem bestimmten Rechtsraum.

Um das Zusammenspiel der einzelnen Entitäten besser veranschaulichen zu können, folgt eine grafische Darstellung des Datenmodells.



*PREMIS Datenmodell in Version 2.0*¹⁰

¹⁰ <http://www.loc.gov/premis/v2/premis-2-0.pdf>

6.4 LMER

Tobias Steinke

Die Langzeitarchivierungsmetadaten für elektronische Ressourcen (LMER) wurden von der Deutschen Bibliothek entwickelt. Das Objektmodell basiert auf dem "Preservation Metadata: Metadata Implementation Schema" der Nationalbibliothek von Neuseeland (2003).

Ziele von LMER sind:

- Ergänzung zu existierenden bibliographischen Metadaten, deshalb nur Beschreibung der technischen Informationen zu einem Objekt und der technischen Veränderungshistorie
- Praxisrelevante Beschränkung auf Angaben, die größtenteils automatisch generiert werden können
- Identifizierung der Kernelemente, die für alle Dateikategorien und jedes Dateiformat gültig sind, sowie ein flexibler Teil für spezifische Metadaten
- Abzubilden als XML-Schema
- Dateiformatidentifikation über Referenz zu einer zu schaffenden File-Format-Registry
- Modularer Aufbau zur Integration in Containerformate wie METS

Historie

LMER entstand 2003 aus dem Bedarf für technische Metadaten im Vorhaben LZA-RegBib. Die erste Version 1.0 wurde 2004 als Referenzbeschreibung und XML-Schema veröffentlicht. 2005 erschien eine überarbeitete Version 1.2, die auch Grundlage für die Verwendung im Projekt kopal ist. Die Version 1.2 führte eine starke Modularisierung und damit einhergehende Aufteilung in mehrere XML-Schemas ein, die eine bessere Einbindung in METS ermöglichte. Als Resultat entstand das METS-Profile-Universelles-Objektformat (UOF), das auf METS 1.4 und LMER 1.2 basiert.

Objektmodell

In LMER meint ein Objekt eine logische Einheit, die aus beliebig vielen Dateien bestehen kann. Es gibt einen Metadatenabschnitt zum Objekt und je einen Metadatenabschnitt zu jeder zugehörigen Datei. Zum Objekt einer jeden Datei kann es Prozess-Abschnitte geben. Diese beschreiben die technische Veränderungshistorie, also vor allem die Anwendung der Langzeiterhaltungsstrategie Migration. Schließlich gibt es noch den Abschnitt Metadatenmodifikation, der

Änderungen an den Metadaten selbst dokumentiert und sich auf alle anderen Abschnitte bezieht. Dabei wird davon ausgegangen, dass sich alle relevanten Metadatenabschnitte in derselben XML-Datei befinden.

Die vier möglichen Abschnittsarten LMER-Objekt, LMER-Datei, LMER-Prozess und LMER-Modifikation werden jeweils durch ein eigenes XML-Schema beschrieben. Dadurch kann jeder Abschnitt eigenständig in anderen XML-Schemas wie METS eingesetzt werden. Es gibt jedoch auch ein zusammenfassendes XML-Schema für LMER, das anders als die einzelnen Schemas Abhängigkeiten und Muss-Felder definiert.

LMER-Objekt

Die Metadaten zum Objekt stellen über einen Persistent Identifier den Bezug zu bibliographischen Metadaten her. Zugleich finden sich dort u.a. Informationen zur Objektversion und zur Anzahl der zugehörigen Dateien.

LMER-Datei

Zu jeder Datei werden die technischen Informationen erfasst, wie sie auch von einem Dateisystem angezeigt werden (Name, Pfad, Größe, Erstellungsdatum), aber auch eine Referenz zu exakten Formatbestimmung. Zudem wird jede Datei einer Kategorie zugeordnet (Bild, Video, Audio etc.), die insbesondere für die spezifischen Metadaten relevant ist. Denn in einem speziellen Platzhalterelement des Datei-Abschnitts können dank des flexiblen Mechanismus von XML-Schemata beliebige XML-Metadaten zur spezifischen Bestimmung bestimmter Dateicharakteristiken hinterlegt werden. Ein Beispiel dafür ist die Ausgabe des Dateianalysewerkzeugs JHOVE.

LMER-Prozess

Die Metadaten in einem Prozess-Abschnitt beschreiben die Schritte und Resultate von technischen Veränderungen und Konvertierungen (Migrationen) an einem Objekt oder einzelnen Dateien eines Objekts. Gehört ein Prozess-Abschnitt zu einem Objekt, so bezeichnet er auch die Versionsnummer und die Kennung des Objekts, von dem die vorliegende Version abgeleitet wurde.

LMER-Modifikation

Die LMER-Daten werden in der Regel in einer oder mehreren XML-Dateien gespeichert. Veränderungen (Ergänzungen oder Korrekturen) der XML-Daten darin können im Modifikationsabschnitt aufgeführt werden.

Literatur

Referenzbeschreibung zu LMER 1.2:

<http://nbn-resolving.de/?urn=urn:nbn:de:1111-2005041102>

Referenzbeschreibung zum Universellen Objektformat (UOF):

http://kopal.langzeitarchivierung.de/downloads/kopal_Universelles_Objektformat.pdf

6.5 MIX

Tobias Steinke

MIX steht für „NISO Metadata for Images in XML“ und ist ein XML-Schema für technische Metadaten zur Verwaltung digitaler Bildsammlungen. Die Metadatenelemente dieses XML-Schemas werden durch den Standard ANSI/NISO Z39.87-2006 („Technical Metadata for Digital Still Images“) beschrieben. MIX wurde von der Library of Congress und dem MARC Standards Office entwickelt. Neben allgemeinen Informationen zu einer Datei werden insbesondere komplexe Informationen zu Bildeigenschaften wie Farbinformationen aufgenommen, sowie detaillierte Beschreibungen der technischen Werte der Erzeugungsgeräte wie Scanner oder Digitalkamera. Zusätzlich kann eine Veränderungshistorie in den Metadaten aufgeführt werden, wobei dies ausdrücklich als einfacher Ersatz für Institutionen gedacht ist, welche keine eigenen Langzeitarchivierungsmetadaten wie PREMIS nutzen. Es gibt keine Strukturinformationen in MIX, denn hierfür wird das ebenfalls von der Library of Congress stammende METS vorgesehen. Die aktuelle Version von MIX ist 1.0 von 2006. Ein öffentlicher Entwurf für MIX 2.0 liegt vor.

Offizielle Webseite: <http://www.loc.gov/standards/mix/>