

# Premis Handreichung

von Thomas Antunes, Annette Handrich, Rainer Heid, Vladimir Marek;  
für den VdW-Arbeitskreis „Elektronische Archivierung“  
September 2011<sup>1</sup>

## 1. Vorbemerkung

Zielsetzung unserer Handreichung ist es, Archiven einen kurzen Abriss darüber zu geben, welchen Nutzen der Einsatz von PREMIS (Preservation Metadata Implementation Strategies) bei der digitalen Archivierung leisten kann. Für eine erste Annäherung an PREMIS wird außerdem ein knapper Überblick über Struktur, Inhalt und Funktionsumfang von PREMIS gegeben. Für eine intensivere Beschäftigung mit PREMIS sei in jedem Fall auf die Auswahlbibliografie im Anhang verwiesen. Auf dieser Basis gibt die Handreichung ein übersichtliches Mind Set für Anforderungen, die sich hieraus an ein elektronisches Archivsystem<sup>2</sup> ergeben, bzw. welche Elemente aus PREMIS im Archivbereich einsetzbar sind.

Premis wurde ursprünglich als Standard für den Bibliotheksbereich entwickelt. Da Premis nicht alle Metadatenbereiche abdeckt, ist eine ausschließliche Ausrichtung eines elektronischen Archivsystems auf PREMIS nicht möglich und muss in jedem Fall durch den Einsatz anderer verfügbarer Standards ergänzt werden (siehe hierzu Punkt 2.3).

## 2. Einführung in PREMIS

### *2.1. Was ist PREMIS?*

- Entstehung: Das „PREMIS Data Dictionary for Preservation Metadata“ wurde 2003-2005 von einer internationalen Arbeitsgruppe von Bibliotheken erarbeitet. Die zweite, derzeit aktuelle Version (Ver. 2.1) wurde im Januar 2011 erstellt. Mit Unterstützung der Library of Congress treibt eine Arbeitsgruppe die Weiterentwicklung von PREMIS voran und stellt mit einer eigenen Website<sup>3</sup> ein wertvolles Informationsinstrument für alle zur Verfügung, die sich mit PREMIS beschäftigen.
- Inhalt: PREMIS liefert zum einen ein eigenes **Datenmodell** zur Abbildung digitaler Objekte und zum anderen einen Katalog von Metadatenelementen - das **Datenwörterbuch** (Data Dictionary)<sup>4</sup> - zur Erhaltung dieser Objekte. Das Datenmodell ist wichtig für das Verständnis des Datenwörterbuchs.

---

<sup>1</sup> Die Autoren danken Frau Dr. Ulrike Gutzmann, Herrn Dr. Christian Keitel und Frau Nicole Sachmann, M.A., für zahlreiche Anregungen und eine kritische Durchsicht des vorliegenden Textes.

<sup>2</sup> Ein elektronisches Archivsystem setzt alle Aspekte der digitalen Archivierung um: Übernahme, Aufbereitung, Erschließung, Retrieval, Datenaufbewahrung und –sicherung.

<sup>3</sup> <http://www.loc.gov/standards/premis/>

<sup>4</sup> Da die meisten Dokumente zu PREMIS in englischer Sprache vorliegen, werden alle PREMIS betreffenden fachterminologischen Begriffe bei ihrer ersten Nennung im Dokument in englischer Sprache in Klammern hinzugefügt.

- Zielsetzung von PREMIS: Es kann als Werkzeug für digitale Archive einen entscheidenden Beitrag leisten zum/zur:
  - Sicherung der Authentizität;
  - Realisierung eines „Trusted Archive“;
  - Qualitätssicherung, z.B. bei der Migration von Daten;
  - Revisions sicheren Archivierung;
  - Gesicherten Austausch von Daten zwischen Archiven zur Vereinfachung von Schnittstellen.

## 2.2. Das Datenmodell von PREMIS und der Begriff der Repräsentation

PREMIS stellt ein Datenmodell zur Verfügung, das neben der Dokumentation der Archivalien zur Erhaltung der signifikanten Eigenschaften<sup>5</sup> auch alle weiteren Hauptprozesse der Datenerhaltung in den folgenden fünf Einheiten, sog. Entitäten (Entities), definiert:

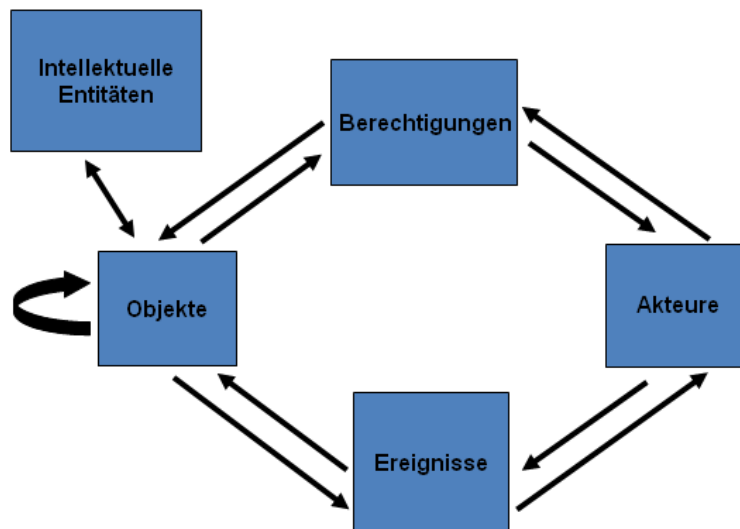
- Intellektuelle Entitäten (Intellectual Entities): „Eine Menge von zusammengehörigen Inhalten, die zu Zwecken des Managements und der Beschreibung als eine einzige intellektuelle Einheit betrachtet wird.“<sup>6</sup> Sie ist die für den Benutzer konzeptionelle Einheit, z.B. ein Buch, eine Fotografie oder eine Datenbank unabhängig von deren physischer Repräsentation. Eine intellektuelle Entität kann andere intellektuelle Entitäten beinhalten, z.B.: Ein Webaufttritt beinhaltet mehrere Unterkapitel; ein Fotoalbum beinhaltet mehrere Fotos. Das Archiv bestimmt – wie in der analogen Welt – bis auf welche Hierarchiestufen bzw. Ebenen, es seine Inhalte beschreiben will (also das ganze Fotoalbum und / oder dessen einzelne Fotos).
- Objekte (Object Entities): Das gespeicherte Objekt (in drei Ausprägungen)<sup>7</sup>
- Ereignisse (Event Entities): Aktionen und das Management der Digitalen Daten und ihre Erhaltung betreffend, z.B. Migration in andere Formate, Digitalisierungen; d.h. alle Ereignisse, die während des gesamten Lebenszyklus digitaler Daten eintreten können.
- Akteure (Agents Entities): Die Personen, Institutionen (das Archiv) und Programme / Anwendungen, die die Ereignisse durchführen und auf die Daten einwirken.
- Berechtigungen (Rights Entities): Berechtigungen wie z.B. Weisungen, aber auch rechtliche Grundlagen, wie z.B. Copyrights und Lizenzen sowie die daraus ableitbaren Rechte, die die entsprechenden Akteure haben, bestimmte Ereignisse durchzuführen, beispielsweise Sperrvermerke.

---

<sup>5</sup> Die Definition der Signifikanz orientiert sich hier an der Ausarbeitung „Eigenschaften digitaler Objekte (2009)“ des VdW - Arbeitskreises "Elektronische Archivierung", vgl. hierzu unter <http://www.wirtschaftsarchive.de/akea/handreichung.htm>, aufgerufen am 07.04.2011.

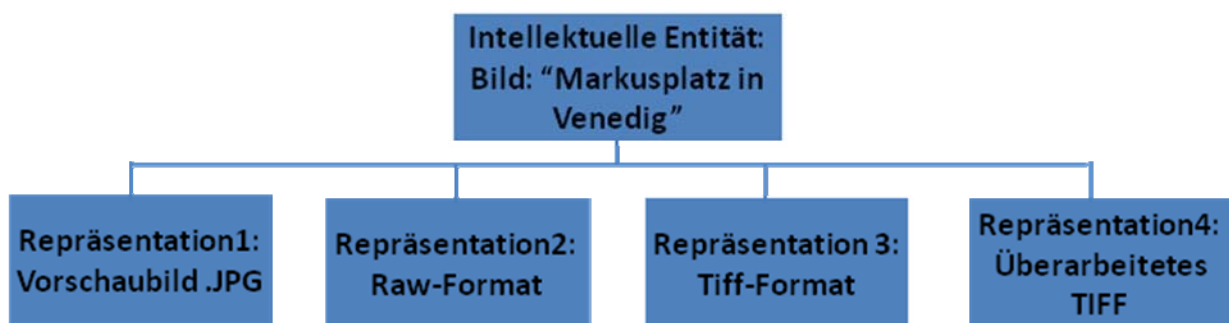
<sup>6</sup> S. Caplan, PREMIS verstehen, [http://www.loc.gov/standards/premis/understanding\\_premis\\_german.pdf](http://www.loc.gov/standards/premis/understanding_premis_german.pdf), S. 9, aufgerufen am 07.04.2011.

<sup>7</sup> Unter diesen Ausprägungen wird verstanden: a) (einzelne) Datei (z.B. PDF); b) representation (z.B. alle Dateien, die eine elektronische Akte bilden); c) bitstream (Teilmengen von Dateien, belegen die Authentizität und Integrität der Datei)



Der neuartige Ansatz des PREMIS-Datenmodells ermöglicht die Trennung der logischen Ebene von der physischen Ebene. Eine intellektuelle Entität (Intellectual Entity), als logisches Objekt verstanden, kann in verschiedenen physischen Formen dieses Objektes (Object Entity) abgebildet und dokumentiert werden. Diese werden als Repräsentationen bezeichnet. Eine Repräsentation als physisches Objekt bildet den Informationskern eines logischen Objekts vollständig ab. Man kann die Verzeichnungseinheit, zu der eine Titelaufnahme gebildet wird, als intellektuelle Entität verstehen. Zu einer Titelaufnahme, etwa ein Film, können verschiedenen Repräsentationen der Archivalie existieren, z.B. ein 8-mm-Film, ein VHS-Band und eine DVD.

#### Repräsentation in PREMIS



#### 2.3. Das PREMIS-Datenwörterbuch und Hinweise zu seiner Umsetzung

- PREMIS liefert mit dem Datenwörterbuch (Data Dictionary) ein Kernset an Metadatenelementen – den so genannten semantischen Einheiten (Semantic Units). Premis ist als Vorschlagskatalog zu verstehen, dessen Umsetzung von Archiv zu Archiv je nach Anforderung sehr unterschiedlich aussehen kann.

<sup>8</sup> Schaubild nach PREMIS Data Dictionary for Preservation Metadata Version 2.1, S. 5.

- Die Metadatenelemente sind im PREMIS-Standard als Eigenschaften der bereits in 2.2 genannten Entitäten definiert. Es handelt sich um die Eigenschaften von:
  - Objekten (Object Entities);
  - Ereignissen (Event Entities);
  - Akteuren (Agent Entities);
  - Berechtigungen (Rights Entities).

Die meisten Metadatenelemente dienen dabei der Beschreibung der Objekte (= Object Entity), dies kann sein: Datei, Repräsentation, Bitstream.

- Für einige Metadatenbereiche stellt PREMIS bewusst wenige bzw. gar keine Metadatenelemente zur Verfügung, wie z.B.
  - Spezifische technische formatspezifische Metadaten.
  - Inhaltliche, die Archivalie beschreibende Metadaten („Erschließungsmetadaten“)
  - Metadaten zu Berechtigungen, die die Abgabe, Verleih etc. von Archivalien nach außen zum Kunden hin regeln.

Hier müssen die fehlenden Angaben anderen Standards entnommen werden, z.B. den Standards für inhaltliche Metadaten. Im Bibliotheksbereich sind dies etwa Marc21, Dublin Core und auch dem ISO Standard 23081 (Teile 1 und 2). Für Archive kommt ebenfalls der Dublin Core-Standard, aber auch z.B. EAD in Frage oder diese Standards müssen selbst definiert werden (z.B. formatspezifische Metadatenelemente). Im Premis Datenwörterbuch werden diese „externen Metadaten“ über sogenannte Erweiterungscontainer (Extensions) implementiert.<sup>9</sup>

- Softwareentwickler können das PREMIS Datenwörterbuch als einen Leitfaden dafür einsetzen:
  - welche Informationen von der Software erfasst und aufgezeichnet werden sollten und wie diese zu strukturieren sind.
  - welche Informationen im Rahmen des Archivmanagements an übergeordneter Stelle dargestellt werden. (z.B. wenn das Handbuch des Archivs festlegt, dass digitale Objekte im PDF-Format abzulegen sind: Dies muss dann nicht mehr explizit für jedes einzelne Objekt hinterlegt werden).
- Ein Archiv muss damit für seine Organisation entscheiden, welche Metadatenelemente aus dem PREMIS-Datenwörterbuch für die elektronische Archivierung benötigt werden. Es muss dabei aufgrund der Anforderungen auch über die „Archivierungstiefe“ entscheiden, d.h. auf wie vielen physischen Ebenen Objekte verwaltet werden sollen:
  - Dateiebene,
  - Repräsentationsebene,
  - Bitstreamebene.
- Eine Repräsentation kann mehrere Dateien umfassen, die sich ihrerseits aus Bitstreams zusammensetzen. Metadaten zu den einzelnen Dateien sind möglich. Am Beispiel des logischen Objektes „Fotoalbum“ umfasst die Repräsentationsebene alle Dateien, die das logische Objekt „repräsentieren“ – dies können z.B. sein:
  - das gesamte digitale Fotoalbum = Eine PDF-Datei

---

<sup>9</sup> Eine aktuelle und kurze Übersicht zur Metadatierung im Records Management mit weiterführenden Literaturhinweisen siehe: Peter Toebak, Records Management – Gestaltung und Umsetzung, Baden 2010, S. 171-180.

- und/oder einzelne Kapitel dieses Fotoalbums = Eine PDF-Datei pro Kapitel
- und/oder die einzelnen digitalen Seiten des Fotoalbums = Eine PDF-Datei pro Seite.

Eine PDF-Datei setzt sich wiederum aus mehreren Bitstreams zusammen.

- PREMIS spezifiziert nicht, wie die Metadaten in einem System repräsentiert werden sollen, sondern definiert ausschließlich Inhalte und Struktur dessen, was das System wissen soll und was es in andere Systeme exportieren können soll.

#### 2.4. PREMIS – nur für digitale Unterlagen?

- Der PREMIS-Standard (sowohl Datenmodell als auch Datenwörterbuch) bezieht sich ausschließlich auf die Archivierung digitaler Objekte/Unterlagen. Da in Archiven häufig Archivalien vorhanden sind, die sowohl digitale als auch analoge Repräsentationsformen haben (z.B. Pläne), kann der Standard pragmatisch auf die Beschreibung analoger Objekte ausgeweitet werden (Modell des Landesarchivs Baden-Württemberg)<sup>10</sup>. Ziel ist dabei, ein Nachweissystem für alle Archivalien zu haben, unabhängig von ihrem Format.

#### 2.5. Anknüpfungspunkte zur Nutzung von PREMIS im Zusammenhang mit METS und OAIS

- Das PREMIS Datenwörterbuch ist mit XML kompatibel. Die semantischen Einheiten von PREMIS können als XML-Elemente implementiert werden. Damit wird der Export nach sowie der Import aus anderen Metadatenschemata, wie z.B. Metadata Encoding and Transmission Standard (METS) ermöglicht. PREMIS kann durch die XML-Kompatibilität bei der Automatisierung von Archivprozessen unterstützend wirken.
- Im Rahmen des OAIS (Open Archival Information System) kann PREMIS für die Erstellung der Informationspakete genutzt werden.<sup>11</sup> Der Standard wirkt beim Zusammenbau der Container. Die Primärdaten der Informationspakete bestehen aus dem eigentlichen digitalen Objekt und seinen Repräsentationsinformationen. Das digitale Objekt wiederum setzt sich laut OAIS aus einer Reihe von Bitsequenzen zusammen. Die Definition des Informationspaketes kann durch das Datenmodell von PREMIS erweitert werden, indem zwischen logischer und physischer Ausprägung eines digitalen Objekts unterschieden wird.<sup>12</sup> So kann man zwischen den verschiedenen Repräsentationen eines Objekts unterscheiden.

---

<sup>10</sup> Vgl. dazu die Vorstellung des DIMAG durch Dr. Christian Keitel anlässlich der 12. Tagung des VdW-Arbeitskreises "Elektronische Archivierung".

<sup>11</sup> Als Informationspakete werden in OAIS logische Container, die die Primärdaten und seine Archivierungsmetadaten (PDI=Preservation Description Information) enthalten, verstanden (SIP=Submission Information Package / Übernahmepaket, AIP=Archival Information Package / Archivpaket, DIP=Dissemination Information Package / Nutzungspaket).

<sup>12</sup> Allerdings sind die „Repräsentation“ von PREMIS und die „Representation Information“ von OAIS ebenso wie die Datenmodelle von Premis und OAIS zu unterscheiden, sie sind nicht identisch. Vgl. als Beispiel für die Anwendung von OAIS und PREMIS: Nicole Sachmann, Das OAIS-Referenzmodell in der Praxis: Ein Konzept zur Archivierung des BR-Intranets, Abschlussarbeit am IID Potsdam, 2010.

### 3. Anforderungen an Struktur und Funktionalitäten eines elektronischen Archivsystems

Die untenstehenden Anforderungen an ein elektronisches Archivsystem basieren auf der vorläufigen Annahme, dass die gängigen auf dem Markt zur Verfügung stehenden Produkte PREMIS bislang nicht oder nur teilweise unterstützen.

Die Anforderungen im Einzelnen:

- Bei Archivalien muss die logische Ebene von der physischen Ebene getrennt werden können.
- Eine intellektuelle Einheit (Intellectual Entity) lässt sich beliebig vielen Repräsentationen zuordnen. Jedes logische Objekt muss mindestens eine Repräsentation besitzen.
- Logische Objekte, physisch gespeicherte Objekte (Object Entity), Ereignisse, Akteure und Berechtigungen müssen separat dokumentierbar sein. Ihnen müssen die entsprechenden Metadatenelemente zugeordnet werden können.
- Automatisierte Übernahme und Zuordnung von Metadaten: Die meisten der PREMIS-Elemente sind so gestaltet, dass sie im Prinzip in einem Archivierungssystem automatisch übernommen werden können. Das heißt jedoch nicht, dass sie von derzeit verfügbarer Software auch bereitgestellt werden. Die Software ist dahingehend zu prüfen und es ist dabei auf einen möglichst hohen Automatisierungsgrad zu achten.
- Abbildbarkeit der Struktur der Premis-Metadaten (Hierarchische Darstellung der Metadaten):
  - Container dienen zur Gruppierung von Metadatenelementen, nur letztere haben einen eigenständigen Wert.
  - Erweiterungscontainer dienen zur Ergänzung von Metadatenelementen, die nicht in PREMIS Datenwörterbuch enthalten sind (z.B. Erschließungsmetadaten).
  - Containerstrukturen müssen in der Regel beliebig wiederholbar sein. Beispiel: Ein physisches gespeichertes Objekt muss mit beliebig vielen Ereignissen in Beziehung gesetzt werden können.
- Das zum Einsatz kommende Archivsystem muss skalierbar sein bzw. modular ausbaubar sein; d.h. Metadatenelemente, die ursprünglich nicht direkt im Archivsystem erfasst werden sollten, müssen nachträglich strukturell integriert werden können (z.B. bei Datenaustausch).
- In PREMIS ist jedem logischen Objekt, jedem physisch gespeicherten Objekt, jedem Akteur, jeder Berechtigung und jedem Ereignis ein Identifikator (identifier) zugeordnet. Dies kann folgendermaßen aussehen: Das „Ereignis Migration nach PDF/A“ hat die Nr. 4, das Ereignis „Ausleihe“ hat die Nr. 17. Die eingesetzten Akteure haben ebenfalls feststehende Nummern, durch die Identifikatoren lassen sich Beziehungen/Verknüpfungen zwischen den PREMIS Entitäten herstellen. Zum Beispiel wird der Identifikator eines spezifischen Ereignisses in die Metadaten des physisch neu gespeicherten Objektes aufgenommen. Z.B. Ereignis 4 durch Akteur 3 am TT.MM.JJJJ durchgeführt. Vorteil dabei ist, dass spezifische Ereignisse, Akteure etc. nur einmal dokumentiert werden

müssen. Der Identifikator für Ereignis 4 kann mit beliebig vielen Objekten verknüpft werden.

- Das Archivsystem muss damit umgehen können und z.B. erkennen, welcher Beziehungstyp nach bestimmten erfolgten Ereignissen vorliegt (z.B. strukturelle Beziehung: „A ist Teil von B“ oder abgeleitete Beziehung: „B ist eine Version von A“). Verbindungen können existieren zwischen:
  - Logischen Objekten untereinander (z.B. kann ein Webaufttritt mehrere Unterkapitel beinhalten)
  - Logischen Objekten und physischen Objekten (z.B. kann eine Bildinformation eine Beziehung zu einem digitalen Bild haben)
  - Physische gespeicherte Objekten untereinander (z.B. zwischen den zu einer Webseite gehörenden Dateien)
  - Physischen Objekten und Ereignissen
  - Physischen Objekten und Berechtigungen
  - Ereignissen und Akteuren
  - Akteuren und Berechtigungen
- 1:1-Prinzip: jedes physisch gespeicherte Objekt (Repräsentation, Datei oder Bitstream) muss dauerhaft und unverändert erhalten bleiben (ein statisches Gebilde von Bits oder Bytes). Eine bestehende Datei darf damit nicht verändert werden, es kann nur eine neue Datei erstellt werden. Daher sieht das PREMIS-Datenwörterbuch nur ein Metadatenelement für die Dateierstellung, nicht jedoch für eine Dateiänderung vor. Die Metadatenelemente der neuen Datei verweisen auf das Ereignis, das zur Erstellung der neuen Datei geführt hat. Damit dient PREMIS auch einer erhöhten Nachvollziehbarkeit aller Archivierungsprozesse, da es die Entwicklung von Objekten darstellen kann.
- Das hier dargestellte Mindset stellt keineswegs einen kompletten Anforderungskatalog dar, er orientiert sich speziell am Metadaten-Schema von PREMIS. Weitere Anforderungen an das Archivsystem ergeben sich aus anderen Standards!

#### 4. Fazit

- Bei unserer Recherche zur Anwendung von PREMIS in der Archivpraxis haben wir den Eindruck gewonnen, dass die uns bekannten Anwender das Prinzip der Trennung von logischer und physischer Einheit, d.h. Elemente des Datenmodells übernommen haben, ohne gleichzeitig die komplexe Struktur des Datenwörterbuchs angewendet zu haben.<sup>13</sup>
- Das PREMIS-Konzept der Repräsentation, d.h. der Idee, dass ein Archivobjekt mehrere inhaltsgleiche Erscheinungsformen haben kann, hat sich in der digitalen Bestandserhaltung durchgesetzt. Gleichzeitig versuchen die Architekten digitaler Archive,

---

<sup>13</sup> Siehe z.B. die Erfahrungen in Baden-Württemberg: <http://www.landesarchiv-bw.de/web/44348>, aufgerufen am 12.5.2011, sowie in Christian Keitel, Das Repräsentationenmodell des Landesarchivs Baden-Württemberg, in: Susanne Wolf (Hrsg.), Neue Entwicklungen und Erfahrungen im Bereich der digitalen Archivierung: von der Behördenberatung zum Digitalen Archiv. 14. Tagung des Arbeitskreises „Archivierung von Unterlagen aus digitalen Systemen“, München 2010, S. 69 – 82 sowie dem Fachkonzept zum Elektronischen Archiv Staatsarchiv Sachsen, [http://www.archiv.sachsen.de/download/100927\\_KO\\_LeA\\_Fach\\_ELArch\\_WEB.pdf](http://www.archiv.sachsen.de/download/100927_KO_LeA_Fach_ELArch_WEB.pdf), S. 18 und S. 94, aufgerufen am 12.5.2011

den Aufwand für Metadaten möglichst gering zu halten, da komplexe Metadatenstrukturen den Bearbeitungsaufwand erhöhen und eine spätere Migration erschweren.<sup>14</sup>

- In PREMIS ist die Strukturierung und Dokumentation von Metadatenelementen im Rahmen der digitalen Datenarchivierung und -erhaltung umfangreich, jedoch sieht PREMIS von seinem Ansatz her keinesfalls die zwingende Übernahme sämtlicher Konzeptpunkte vor.<sup>15</sup> Dies bestätigen die uns bekannten Beispiele der tatsächlichen Anwendung.
- Grundsätzlich geht es darum, das für die Archivierung notwendige Kernset an Metadaten aus Premis aufzunehmen und dieses durch die aus anderen Standards für die Archivierung notwendigen Metadaten zu ergänzen. Man kann von PREMIS auch nur Teile nutzen, die für die eigene Umgebung sinnvoll erscheinen, man muss nicht alle Aspekte 1:1 umsetzen.
- PREMIS ist schwere Kost und erfordert in jedem Fall vor der Implementierung ein vergleichsweise hohes Maß an Einarbeitung in die theoretischen Grundlagen.

## 5. Auswahlbibliografie

- PREMIS Data Dictionary for Preservation Metadata, version 2.1., 2011  
<http://www.loc.gov/standards/premis/v2/premis-2-1.pdf>, aufgerufen am 12.5.2011
- PREMIS Tutorial, von Rebecca Guenther u. Angela Dappert, 2009  
<http://www.loc.gov/standards/premis/>  
→ dann unter Tutorials, Part 1, PREMIS Introduction and Data Model
- Priscilla Caplan: Understanding PREMIS, 2009  
<http://www.loc.gov/standards/premis/understanding-premis.pdf>, aufgerufen am 12.5.2011
- Deutsche Version verfügbar unter:  
[http://www.loc.gov/standards/premis/understanding\\_premis\\_german.pdf](http://www.loc.gov/standards/premis/understanding_premis_german.pdf), aufgerufen am 12.5.2011
- Olaf Brandt:  
[http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor\\_handbuch\\_artikel\\_357.pdf](http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_357.pdf),  
aufgerufen am 12.5.2011  
[http://www.iwiiuk.org/cashmere/htdocs/html/workshop/presentations/brandt\\_premis.pdf](http://www.iwiiuk.org/cashmere/htdocs/html/workshop/presentations/brandt_premis.pdf),  
aufgerufen am 3.2.2010

---

<sup>14</sup> Vgl. z.B. [http://www.landesarchiv-bw.de/sixcms/media.php/25/konzeption\\_metadaten10.pdf](http://www.landesarchiv-bw.de/sixcms/media.php/25/konzeption_metadaten10.pdf), aufgerufen am 07.04.2011.

<sup>15</sup> Vgl. PREMIS-Standard V. 2.1 S. 15,16; <http://www.loc.gov/standards/premis/premis-conformance-oct2010.pdf>, aufgerufen am 07.04.2011.