

H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, K. Huth (Hrsg.)

nestor Handbuch

Eine kleine Enzyklopädie
der digitalen Langzeitarchivierung

Version 2.3

Kapitel 6.3

PREMIS

nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung
hg. v. H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, K. Huth
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

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.

Neben der Online Version 2.3 ist eine Printversion 2.0 beim Verlag Werner Hülsbusch,
Boizenburg erschienen.

Die digitale Version 2.3 steht unter folgender Creative-Commons-Lizenz:
„Namensnennung-Keine kommerzielle Nutzung-Weitergabe unter gleichen Bedingungen 3.0
Deutschland“
<http://creativecommons.org/licenses/by-nc-sa/3.0/de/>



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.

URL für Kapitel 6.3 „PREMIS“ (Version 2.3): [urn:nbn:de:0008-2010061796](http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-2010061796)
<http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-2010061796>



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

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/activities/past/orprojects/pmwg/pm_framework.pdf

3 Abschlußbericht der PREMIS Arbeitsgruppe mit „Data Dictionary for Preservation Metadata“: <http://www.oclc.org/research/activities/past/orprojects/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

Entities sind abstrakte Klassen von 'Dingen', also z.B. „digitale Objekte“ oder

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/standards/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>

„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.

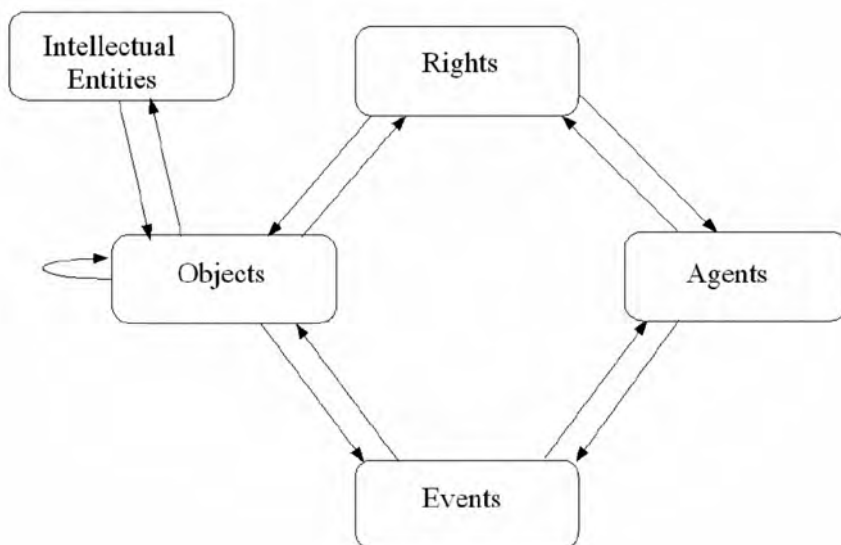
Agent Entity

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>