

PREMIS

Aktuelle Entwicklungen und Perspektiven

Olaf Brandt



Gliederung

- Hintergrundinformationen zu PREMIS
- Aktivitäten
- Welchen Stand haben wir erreicht?
- Wohin geht die Entwicklung?



Was ist PREMIS?

- Das Akronym steht für PREservation Metadata: Implementation Strategies
- PREMIS wurde 2003 mit Unterstützung der Research Libraries Group (RLG) und des Online Computer Library Center (OCLC) gegründet
- Die Initiative ist ein internationaler Verbund von Akteuren aus dem Umfeld von Gedächtnisorganisationen (Archiven, Bibliotheken, Museen ...) und der Privatwirtschaft, die im Bereich digitale Langzeitarchivierung tätig sind.
- PREMIS baut auf dem Vorläufer, der Preservation Metadata Working Group (PMWG) auf.

Was macht PREMIS?

- Untersucht alternative Strategien für die Erstellung und Nutzung von Encodings, Speicherung und Management von Langzeitarchivierungsmetadaten
- Untersuchte produktive und geplante Implementierungen von digitalen Langzeitarchiven und dabei auftretende Probleme und Fragen
- liefert 'best practices'
- Erstellte ein in der Community weit anwendbares und implementierbares Kernset an Langzeitarchivierungsmetadaten
- PREMIS betreut initiale Piloten bei der Umsetzung der Empfehlungen in der PREMIS Maintenance Activity

Was sind Langzeitarchivierungsmetadaten?

Es sind Informationen, die helfen, digitale Objekte langfristig verfügbar zu halten. Das sind z.B.:

Provenance Metadaten:

Wem gehörte und wer hielt das digitale Objekt?

Welche Maßnahmen wurden unternommen, um das digitale Objekt verfügbar zu halten?

Authentizität:

Ist das digitale Objekt das, was es zu sein scheint?

Technisches Environment:

Aus welcher technischen Umgebung stammt das digitale Objekt und was ist notwendig, um es darzustellen und erlebbar zu machen?

Rechte-Metadaten:

Welche Rechte müssen beachtet werden?

PREMIS Data Dictionary

- Im Mai 2005 wurde das PREMIS Data Dictionary publiziert und findet weite Beachtung (über 4600 Downloads des kompletten Berichtes und über 2700 Downloads von Teilen des Berichtes - Stand Ende August).
- Der Abschlussbericht der PREMIS Arbeitsgruppe beinhaltet das PREMIS Data Dictionary 1.0, welches von einem ausführlichen Bericht begleitet wird.
- In ihm werden der Kontext, das Datenmodell und die PREMIS Grundannahmen erörtert. Spezielle Themen, ein Glossar und Beispiele sind ebenfalls enthalten.
- Es ist die praktische Grundlage für die Implementierung von den PREMIS Langzeitarchivierungsmetadaten in digitalen Archiven.

Welche Bereiche deckt das PREMIS Data Dictionary ab?

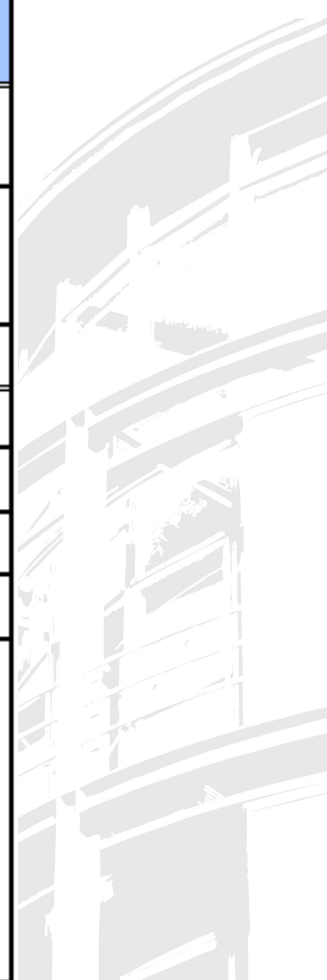
- Es soll implementierungsunabhängig sein
- Deskriptive Metadaten sind nicht Teil der PREMIS-Aktivitäten
- Technische Metadaten sind nur insoweit berücksichtigt, als sie bei allen Datenformaten auftreten
- Medien- oder Hardwareinformationen sind nur in Kernbereichen behandelt
- Business rules sind nicht berücksichtigt worden
- Die Rechte-Metadaten sind nur für die Langzeitarchivierung gedacht, nicht für den Zugriff

Hauptkategorien

- Object Entity
- Event Entity
- Rights Entity
- Agent Entity



Semantic unit	formatRegistry		
Semantic components	formatRegistryName, formatRegistryKey, formatRegistryRole		
Definition	Identifies and/or gives further information about the format by reference to an entry in a format registry.		
Rationale	If central format registries are available to the preservation repository, they may provide an excellent way of referencing detailed format information.		
Data constraint	Container		
Object category	Representation	File	Bitstream
Applicability	Not applicable	Applicable	Applicable
Repeatability		Repeatable	Repeatable
Obligation		Optional	Optional
Usage notes	<p>Either formatDesignation or at least one instance of formatRegistry is required.</p> <p>The PREMIS working group assumed that a number of format registries will be developed and maintained to support digital preservation efforts. The proposal for a Global Digital Format Registry (GDFR), for example, would create a network-accessible registry designed to store detailed specifications on formats and profiles.</p>		



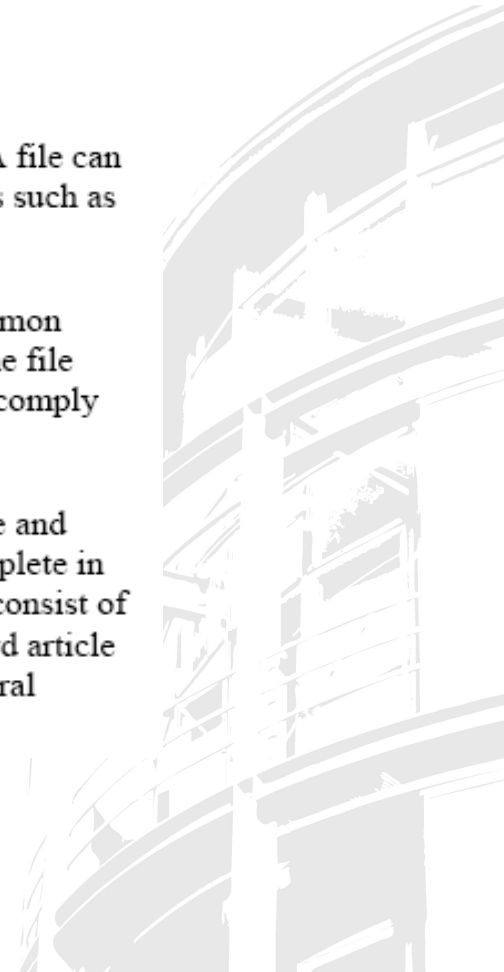
Objects

The Object entity has three subtypes: file, bitstream, and representation.

A **file** is a named and ordered sequence of bytes that is known by an operating system. A file can be zero or more bytes and has a file format, access permissions, and file system statistics such as size and last modification date.

A **bitstream** is contiguous or non-contiguous data within a file that has meaningful common properties for preservation purposes. A bitstream cannot be transformed into a standalone file without the addition of file structure (headers, etc.) and/or reformatting the bitstream to comply with some particular file format.

A **representation** is the set of files, including structural metadata, needed for a complete and reasonable rendition of an Intellectual Entity. For example, a journal article may be complete in one PDF file; this single file constitutes the representation. Another journal article may consist of one SGML file and two image files; these three files constitute the representation. A third article may be represented by one TIFF image for each of 12 pages plus an XML file of structural metadata showing the order of the pages; these 13 files constitute the representation.



Implementierungsphase

- Das Data Dictionary ist für die nächsten 18 Monate stabil und wird erst danach bei Bedarf angepasst. Zur Zeit befindet es sich in einer Test- und Implementierungsphase.
- Es wurde ein Set von XML-Schemata entwickelt.
- Die Schemata sollen in spezifische Kontexte eingebunden und implementiert werden (z.B. in METS).
- Metadaten sollen maschinell erzeugt werden.
- Dazu müssen und werden Tools entwickelt.

Perspektiven und offene Fragen

- Welche Rolle werden Registries (Format- und Environmentregistries) spielen und wie werden sie integriert?
- Ein wichtiger Bereich für PREMIS wird der Bereich des Zusammenarbeit und des Datenaustausches sein.
- Best Practices werden in der Implementors Gruppe erörtert und verbreitet.
- Das PREMIS Metadatenmodell und -set sollte ein Standardisierungsverfahren durchlaufen.
- PREMIS benötigt Unterstützung auf dem weiteren Weg zur Implementierung und Standardisierung.

Sehr gutes Feedback

“Der diesjährige von der Digital Preservation Coalition gesponsorte Digital Preservation Award geht an die PREMIS Arbeitsgruppe für ihre Verdienste um die Etablierung eines internationalen Standards im Bereich der Langzeitarchivierungs-Metadaten. Der Preis wurde anlässlich der Verleihung der Conservation Awards in der British Library verliehen. Das Preisgeld beträgt 5000 £.”

URL: <http://nestor.sub.uni-goettingen.de/aktuell/index.php#143>
am 23.11.2005

URLs

PREMIS Arbeitsgruppe: <http://www.oclc.org/research/projects/pmwg/>

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

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

“Implementing Metadata in Digital Preservation Systems: The PREMIS Activity”. *D-Lib
Magazine* (April 2004) <http://www.dlib.org/dlib/april04/lavoie/04lavoie.html>

Implementing Preservation Repositories for Digital Materials, Current Practice and
Emerging Trends in the Cultural Heritage Community (survey report):

<http://www.oclc.org/research/projects/pmwg/surveyreport.pdf>

International team wins the 2005 Digital Preservation Award:
<http://www.dpconline.org/graphics/advocacy/press/award2005.html>

Preservation Metadata Working Group (PMWG 2002) Framework:
http://www.oclc.org/research/projects/pmwg/pm_framework.pdf

RLG DigiNews Oktober 2004 und Dezember 2004:
http://www.rlg.org/en/page.php?Page_ID=12081

Vielen Dank!

brandt@sub.uni-goettingen.de



Nachfolgend noch einige Definitionen und Beispiele



Agent Entity

Die Agent Entity enthält Informationen über Attribute oder besondere Eigenschaften von Agenten, die im Zusammenhang mit Langzeitarchivierungsereignissen und Rechtemanagement im Leben eines Datenobjektes auftreten. Informationen über Agenten dienen der eindeutigen Identifizierung eines Agenten.

Event Entity

Die Event Entity enthält Informationen über eine Aktion/ein Ereignis, bei dem ein oder mehrere Object Entities einbezogen wurden. Metadaten über ein Event (Ereignis/Aktion) werden normalerweise getrennt vom Datenobjekt aufgenommen und gespeichert.

Ob ein digitales Langzeitarchiv ein Ereignis aufzeichnet, hängt von der Wichtigkeit des Ereignisses ab.

Ereignisse, die das Objekt verändern, sollten immer aufgenommen werden. Andere Ereignisse, wie z. B. ein Backup können auch in Systemlogs o. ä. aufgenommen werden.

Intellectual Entity

Eine zusammenhängende Reihe von Inhalten, die als Einheit beschrieben wird. Z. B. Ein Buch, eine (Land-)Karte ... Eine Intellectual Entity kann andere Intellectual Entities beinhalten. Z. B. Ein gebundener Jahrgang einer Zeitschrift kann mehrere Zeitschriften enthalten, die wiederum einzelne Artikel enthalten können, die wiederum Abbildungen enthalten können usw.

Object Entity

Die Object Entity enthält Informationen für die Langzeitarchivierung über ein Datenobjekt und beschreibt die relevanten Eigenschaften für das Preservation Management.

Das Datenobjekt wird im digitalen Langzeitarchivsystem gehalten und dazu wahrscheinlich passende Metadaten (z. B. deskriptive, technische und administrative) haben, die durchaus in einem externen System gehalten werden können.

Rights Entity

Aussagen über Rechte und Erlaubnisse sind Konstruktionen, die als Rights Entity bezeichnet werden. Rechte sind Ermächtigungen, die Akteure per Copyright oder anderen Gesetzen bekommen. Erlaubnisse sind Macht oder Privilegien die durch Vereinbarung oder Vertrag zwischen einem Rechthehalter und einer oder mehreren anderen Parteien verliehen werden.

Ein Langzeitarchiv könnte den Bedarf haben, mehrere unterschiedliche Rechteinformationen aufzunehmen. Sie könnten abstrakte Aussagen über Rechte, Erlaubnisse für externe Agenten und über Objekte, die nicht im Langzeitarchiv gehalten werden enthalten. Ein Langzeitarchiv sollte mindestens von Kernrechten wissen, die ihm für die tägliche Arbeit mit den Objekten verliehen wurden (kopieren, in ein anderes Format umwandeln etc.).