

Library of Congress Network Development and MARC Standards Office

PREMIS verstehen

Autorin: Priscilla Caplan

Veröffentlichungsdatum: 1. Februar 2009

Copyright © 2009 The Library of Congress, ausgenommen in den U.S.A.
Die Verwendung von Auszügen aus dieser Veröffentlichung ist entsprechend zu kennzeichnen.

Übersetzung: Tobias Beinert, Bayerische Staatsbibliothek

Im Auftrag von nestor – Kompetenznetzwerk Langzeitarchivierung und
Langzeitverfügbarkeit digitaler Ressourcen für Deutschland im Rahmen eines
Projekts des Masterstudiengangs Bibliotheks- und Informationswissenschaft der
Fachhochschule Köln

Inhaltsverzeichnis

1. DER ANWENDUNGSKONTEXT VON PREMIS.....	3
1.1. Was sind Langzeitarchivierungsmetadaten?.....	3
1.2. Was ist PREMIS?.....	4
1.3. Was steht im PREMIS Data Dictionary?.....	4
1.4. Wie soll PREMIS eingesetzt werden?	5
1.5. Sollten Sie PREMIS einsetzen?.....	6
2. KONVENTIONEN DES DATA DICTIONARY.....	7
2.1. Semantische Einheiten.....	7
2.2. Container und untergeordnete Einheiten.....	7
2.3. Erweiterungscontainer.....	8
3. DAS PREMIS DATENMODELL.....	8
3.1. Intellektuelle Entität.....	9
3.2. Entität Objekt.....	9
3.3. Ereignisse	10
3.4. Agenten.....	11
3.5. Rechte	11
4. DAS DATA DICTIONARY.....	12
4.1. Beispieleintrag im Data Dictionary für eine einfache semantische Einheit	12
4.2. Beispieleintrag im Data Dictionary für eine Container-Einheit.....	13
5. DIE ANWENDUNG VON PREMIS.....	15
5.1. PREMIS in XML.....	15
5.2. PREMIS Konformität.....	15
6. WEITERE INFORMATIONEN.....	16
Appendix A: Liste aller semantischen Einheiten in PREMIS.....	18
Appendix B: Beispiel für ein Objekt.....	23
Appendix C: Glossar der Begriffe.....	27

PREMIS VERSTEHEN

Dieser Leitfaden gibt einen relativ knappen Überblick über den Metadatenstandard PREMIS für die digitale Langzeitarchivierung. Er wird Ihnen nicht genug Informationen für eine Implementierung von PREMIS liefern können, aber einen Eindruck davon vermitteln, um was es bei PREMIS geht. Für viele Leser wird dies ausreichen. Für diejenigen, die das PREMIS Data Dictionary (PREMIS Datenlexikon) für Langzeitarchivierungsmetadaten beherrschen müssen, kann dieser Leitfaden einen leichten Einstieg ermöglichen, um sich mit dem umfassenderen Dokument vertraut zu machen.

Im vorliegenden Dokument bezeichnen Begriffe in Verdana die semantischen Einheiten von PREMIS. Die Begriffe, die sich im Glossar finden, sind bei ihrem ersten Erscheinen *kursiv* gesetzt.

1. DER ANWENDUNGSKONTEXT VON PREMIS

1.1. Was sind Langzeitarchivierungsmetadaten?

Wenn Sie in einer Bibliothek oder in einem Archiv arbeiten, stehen die Chancen gut, dass Sie zumindest schon einmal von Metadaten und der Beschreibung von Ressourcen gehört haben. Sie wissen wahrscheinlich, dass sich Metadaten nach ihrem Zweck einteilen lassen: beschreibende Metadaten helfen beim Auffinden und Identifizieren von Ressourcen, administrative Metadaten helfen bei deren Verwaltung und Verfolgung und strukturelle Metadaten geben Auskunft darüber, wie komplexe digitale Objekte zusammengesetzt sind und dementsprechend dargestellt oder andersweitig genutzt werden können. In ähnlicher Weise unterstützen *Langzeitarchivierungsmetadaten* Aktivitäten, die die dauerhafte Nutzbarkeit von digitalen Ressourcen gewährleisten sollen.

Das PREMIS Data Dictionary definiert Langzeitarchivierungsmetadaten als „Informationen, die ein Langzeitarchiv einsetzt um den Prozess der digitalen Langzeitarchivierung zu unterstützen. Wie Langzeitarchivierungsmaßnahmen mithilfe von Metadaten unterstützt werden können, zeigen die folgenden Beispiele:

- Digitale Ressourcen müssen sicher gespeichert werden, sodass sie von niemanden versehentlich (oder absichtlich) verändert werden können. Werden Checksummen in den Metadaten abgelegt, ist es möglich, festzustellen, ob eine Datei innerhalb eines gewissen Zeitraums verändert worden ist.
- Dateien müssen auf Datenträgern gespeichert werden, die von heutigen Computern gelesen werden können. Wenn die Datenträger beschädigt oder obsolet sind (wie z.B. die in den 1970er Jahren gebräuchlichen 8-Zoll-Disketten), kann es schwierig oder unmöglich sein, die darauf enthaltenen Daten wiederherzustellen. Mit Metadaten kann das Management der Datenträger unterstützt werden, indem der Typ und das Alter des Speichermediums sowie der letzte Zeitpunkt der Erneuerung der Dateien aufgezeichnet werden.
- Über längere Zeiträume können auch weit verbreitete Dateiformate obsolet werden, in dem Sinne, dass sie von aktuellen Anwendungen nicht mehr dargestellt werden können. Die für die Langzeitarchivierung Verantwortlichen müssen *Langzeiterhaltungsstrategien* umsetzen, um sicherzustellen, dass die digitalen Ressourcen brauchbar bleiben. Diese Strategien können darin bestehen, ältere Formate in neuere zu überführen (*Migration*) oder die alte Abspielumgebung auf neuerer Hard- und Software nachzubilden (*Emulation*). Sowohl für Migration als auch für Emulation werden Metadaten zum ursprünglichen Dateiformat und zur Unterstützung durch Hardware- und Softwareumgebungen benötigt.
- Langzeitarchivierungsmaßnahmen machen es teilweise nötig, Ressourcen oder ihre Darstellung zu verändern. Das kann dazu führen, dass die Authentizität der Ressource in Frage gestellt wird. Metadaten können durch das Dokumentieren der *digitalen Provenienz* der Ressource – Verarbeitungskette und Historie der autorisierten Veränderungen – die Authentizität unterstützen.

1.2. Was ist PREMIS?

PREMIS steht für "PREservation Metadata: Implementation Strategies", den Namen einer internationalen Arbeitsgruppe, die von 2003-2005 von OCLC (Online Computer Library Center) und RLG (Research Libraries Group) gefördert wurde. Diese Arbeitsgruppe verfasste einen Report mit dem Namen PREMIS Data Dictionary for Preservation Metadata (PREMIS Datenlexikon für Langzeitarchivierungsmetadaten), der sowohl ein Data Dictionary als auch Informationen zu Langzeitarchivierungsmetadaten enthält. Eine überarbeitete zweite Version wurde im März 2008 veröffentlicht. Die Library of Congress unterhält ein Schema für die Darstellung von PREMIS in XML.

Es gibt eine aktive PREMIS Maintenance Activity (Aktivitäten zur Erhaltung und Pflege von PREMIS), die von der Library of Congress unterstützt wird. Diese umfassen eine Website mit Links zu allen Arten von offiziellen und inoffiziellen Informationen zu PREMIS, eine Diskussionsliste und ein Wiki für PREMIS-Anwender sowie das Herausgeberkomitee, das verantwortlich ist für Überarbeitungen des Data Dictionary und der Schemata. Die Maintenance Activity versucht darüber hinaus den Bekanntheitsgrad von PREMIS zu steigern, unterstützt Tutorials zur Anwendung von PREMIS und gibt Studien und Publikationen mit Bezug zu PREMIS, wie diese Einführung, in Auftrag.

Wird allgemein von „PREMIS“ gesprochen, so beziehen sich die Sprecher dabei in der Regel auf das Data Dictionary. Gelegentlich wird aber auch auf das XML Schema, die Arbeitsgruppe oder auf die gesamte Arbeit, inklusive der Maintenance Activity Bezug genommen.

PREMIS Data Dictionary: www.loc.gov/premis/v2/premis-2-0.pdf

PREMIS Website: www.loc.gov/standards/premis/

PREMIS Anwendergruppe Diskussionsliste: listserv.loc.gov/listarch/pig.html

1.3. Was steht im PREMIS Data Dictionary?

Im PREMIS Data Dictionary wird ein Kernset an sogenannten *semantischen Einheiten* (semantic units) (siehe Kapitel 2.1) definiert. Diese sollten einem Langzeitarchiv bekannt sein, damit es seine Archivierungsfunktion erfüllen kann. Wenngleich Art und Umfang der Langzeitarchivierung für die einzelnen digitalen Archive dabei sehr unterschiedlich sein können, so wird sie jedoch auf alle Fälle Maßnahmen umfassen, die sicherstellen, dass digitale Objekte anwendbar (d.h. sie können von einem Datenträger ausgelesen werden) und darstellbar (d.h. sie können von einer Anwendungssoftware angezeigt, abgespielt oder in anderer Weise interpretiert werden) bleiben. Ebenso soll sichergestellt werden, dass die digitalen Objekte im Langzeitarchiv nicht unabsichtlich verändert werden und dass begründete Veränderungen dokumentiert werden.

Das Data Dictionary soll nicht alle möglichen Metadatenelemente für die Langzeitarchivierung festlegen, sondern nur jene, die den meisten Langzeitarchiven in den meisten Fällen bekannt sein sollten. Einige Metadatenkategorien sind ausgeschlossen, sie befinden sich außerhalb des Anwendungsbereichs. Dazu gehören:

- Formatspezifische Metadaten, d.h. Metadaten, die nur ein einziges Dateiformat oder eine Klasse von Formaten, wie Audio, Video oder Vektorgraphiken betreffen.
- Metadaten, die anwendungs- oder geschäftsablaufspezifisch sind, d.h. Metadaten, die die Policy oder Praxis eines individuellen Archivs beschreiben, z. B. in Bezug auf die Zugänglichmachung von digitalen Materialien.
- Deskriptive Metadaten. Obwohl die inhaltliche Beschreibung von Ressourcen natürlich auch aus Sicht der Langzeitarchivierung durchaus relevant ist, gibt es bereits eine Vielzahl an unabhängigen Standards wie MARC21, MODS und Dublin Core, die zu diesem Zweck eingesetzt werden können.
- Detaillierte Informationen über Speichermedien oder Hardware. Auch hier gilt, dass obwohl diese offensichtlich relevant für die Langzeitarchivierung sind, es anderen Communities überlassen war, Metadaten dafür festzulegen.

- Informationen über Agenten (Personen, Organisationen oder Software), die über das zur Identifizierung notwendige Minimum hinausgehen.
- Informationen über Rechte und Genehmigungen, bis auf jene die sich unmittelbar auf die Langzeitarchivierung auswirken.

Stellt man sich alle Metadaten vor, die eine Organisation benötigt, die ein digitales Langzeitarchiv betreibt, kann PREMIS als die Definition eine Teilmenge im Zentrum betrachtet werden. Auf der einen Seite befasst sich PREMIS nicht mit der Auffindbarkeit und dem Zugang und auf der anderen Seite versucht PREMIS nicht, detaillierte formatspezifische Metadaten zu definieren. PREMIS definiert nur die Metadaten, die allgemein und für alle Materialien zur Durchführung von Langzeitarchivierungsmaßnahmen benötigt werden.

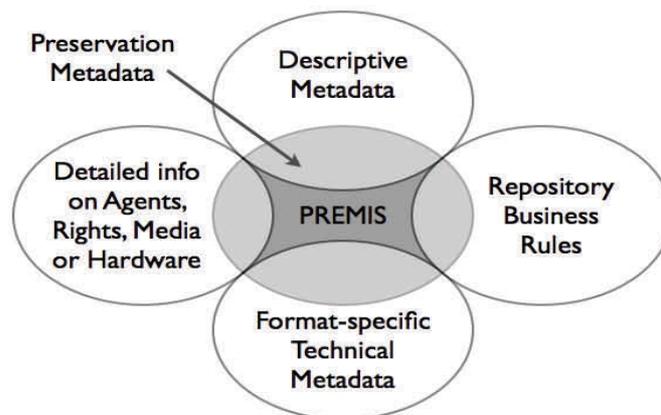


Abbildung 1. PREMIS als eine Teilmenge der gesamten Langzeitarchivierungsmetadaten

Abbildung 1 zeigt alle für die Langzeitarchivierung relevanten Metadaten, als einen grau unterlegten Kreis in der Mitte des Diagramms. Der Kreis enthält einige deskriptive Metadaten, einige Geschäftsabläufe, einige detaillierte technische Metadaten und einige detaillierte Informationen über Agenten, Rechte, Medien und Hardware. PREMIS bildet den kleinen – hier dunkelgrau unterlegten – Kern in der Mitte der Langzeitarchivierungsmetadaten, der all diese Typen ausschließt.

1.4. Wie soll PREMIS eingesetzt werden?

Das PREMIS Data Dictionary definiert, was ein Langzeitarchiv wissen muss. Es ist wichtig festzuhalten, dass der Fokus auf dem Archivsystem und dessen Management liegt und nicht bei den Verfassern von digitalen Inhalten, Personen, die durch Scannen oder andersweitig analoge in digitale Inhalte verwandeln oder auf dem Personal, das kommerzielle elektronische Ressourcen bewertet und lizenziert. Primär kann PREMIS zur Gestaltung und Evaluierung von Langzeitarchiven sowie zum Austausch von archivierten Informationsobjekten zwischen Langzeitarchiven eingesetzt werden.

Diejenigen, die Software für Langzeitarchive gestalten und/oder entwickeln, sollten PREMIS als einen Leitfaden dafür einsetzen, welche Informationen von der Softwareanwendung erfasst und aufgezeichnet werden sollten oder dem Management des Langzeitarchivs in anderer Form bekannt sein sollten.

Diejenigen, die planen, ein Archivsystem zu implementieren, sollten PREMIS als eine Checkliste

einsetzen, um mögliche Softwarelösungen zu bewerten. Systeme, die das PREMIS Data Dictionary unterstützen, werden besser geeignet sein, Informationsressourcen dauerhaft zu bewahren.

Ein aktives Langzeitarchiv wird beizeiten gespeicherte Informationspakete exportieren wollen, um sie in ein anderes Langzeitarchiv zu überführen. So könnte beispielsweise eine treuhänderisch tätige Organisation von einem Archivsystem zu einem anderen migrieren wollen oder ein Kunde möchte von einem externen Serviceanbieter zu einem anderen wechseln. PREMIS stellt dafür ein allgemeines Set von Datenelementen zur Verfügung, das sowohl vom exportierenden als auch vom importierenden Archivsystem verstanden werden kann, insbesondere dann, wenn das PREMIS XML-Schema eingesetzt wird.

1.5. Sollten Sie PREMIS einsetzen?

Es kommt darauf an. Die meisten MitarbeiterInnen in Bibliotheken, Archiven, Museen und anderen kulturellen Gedächtnisorganisationen sind nicht unmittelbar mit der digitalen Langzeitarchivierung befasst. In diesem Fall reicht es aus, zu wissen was PREMIS ist: ein Data Dictionary für Langzeitarchivierungsmetadaten. Wenn Sie aber im Rahmen ihres Jobs Verantwortung für Fragen der Langzeitarchivierung tragen, werden Sie es wahrscheinlich hilfreich finden, mit PREMIS vertraut zu sein. Falls Sie mit der Evaluierung oder Implementierung eines institutionellen Repositoriums oder Archivierungssystems betraut sind, sollten Sie ein umfassendes Verständnis von PREMIS haben. Falls es Ihnen möglich ist, sollten Sie die Teilnahme an einem der regelmäßig angebotenen PREMIS-Tutorials der Maintenance Activity in Betracht ziehen.

Wenn Sie an Digitalisierungsprojekten arbeiten, stellen Sie sich vielleicht die Frage, ob Sie PREMIS-Metadaten für den späteren Gebrauch generieren sollten. Die meisten der PREMIS Elemente sind so gestaltet, dass Sie automatisch von einer Archivierungssystem-Software bereitgestellt werden können. (Auch wenn das natürlich nicht heißt, dass sie von den derzeit verfügbaren Softwareanwendungen bereitgestellt werden). Jedenfalls gibt es einige Informationen, die Sie – falls möglich – aufzeichnen sollten.

Zugriffssperren. Als *Zugriffssperren* sind die Merkmale eines Objekts definiert, die die Zugänglichmachung, die Benutzung oder die Migration verhindern sollen. Zu den Zugriffssperren gehören der Schutz durch Passwörter und die Verschlüsselung. Es ist schwierig, diese Zugriffssperren mit einer Software zu beschreiben, da die Software möglicherweise daran gehindert wird, das Objekt zu analysieren. Deshalb sollten Sie, falls Sie wissen, dass es Zugriffssperren für eine Datei gibt, dies auch festhalten. PREMIS definiert semantische Einheiten für den Typ, das Ziel (welche Handlungen verhindert werden) und den Schlüssel (Passwort oder andere Mechanismen zum Umgehen der Zugriffssperre).

Provenienz. *Digitale Provenienz* ist die Dokumentation der Verarbeitungskette und der Veränderungshistorie eines digitalen Objekts. Wenn Ihre eigene Institution das Objekt erstellt hat, ist der Kontext der Entstehung sicherlich ein wichtiger Teil der Provenienz. Der Name und die Version der zur Erstellung eingesetzten Software kann oft aus dem Datei-Header ausgelesen werden, jedoch nicht in allen Fällen, sodass die Aufzeichnung dieser Informationen empfohlen wird. PREMIS ermöglicht es, die Veränderungshistorie als Ereignis- (Event) Informationen abzulegen – dies ist weiter unten beschrieben. Allerdings sind die PREMIS Ereignistypen in erster Linie dafür gedacht, Handlungen zu beschreiben, die nach der Abgabe von Objekten für die Übernahme in ein Langzeitarchiv durchgeführt werden. Um Ereignisse aufzuzeichnen, die vor der Übernahme stattfinden, wie Auswahl und Erwerbung, müssen Sie möglicherweise Ihre eigenen Ereignistypen schaffen.

Signifikante Eigenschaften. *Signifikante Eigenschaften* sind die Charakteristika eines Objekts, die mittels Langzeitarchivierungsmaßnahmen erhalten werden sollen. Wenn Sie zum Beispiel ein Dokument haben, stellt sich die Frage, ob bloß der Text und die Bilder entscheidend sind oder auch die Schriften, der Hintergrund, die Formatierung und weitere Eigenschaften des „Look and Feel“

eines Objekts genauso wichtig sind? Die Idee von den signifikanten Eigenschaften ist eines der wichtigsten Konzepte in der digitalen Langzeitarchivierung und zugleich eines derer, die am wenigsten verstanden werden. Es gibt eine Reihe von Initiativen, die sich einer besseren Modellierung und Beschreibung der signifikanten Eigenschaften widmen, doch diese befinden sich noch in ihren Anfangsphasen. Dennoch sollten sich alle Institutionen, die für ihre Nutzergruppen digitale Materialien erstellen oder erwerben, gründlich darüber Gedanken machen, welche Eigenschaften dieser Materialien für diese Nutzergruppe von Bedeutung sind und versuchen diese Informationen für den späteren Gebrauch aufzuzeichnen.

Rechte. Rechtliche Informationen sind natürlich nicht ausschließlich für die Langzeitarchivierung relevant, aber zu wissen, was man mit einem Objekt machen darf, ist für den Erhaltungsprozess immens wichtig. Sie sollten deswegen darauf achten, alle bekannten rechtlichen Informationen, inklusive des urheberrechtlichen Status, der Lizenzbedingungen sowie spezielle Befugnisse, aufzuzeichnen.

2. KONVENTIONEN DES DATA DICTIONARY

2.1. Semantische Einheiten

Das PREMIS Data Dictionary definiert keine Metadatenelemente, sondern semantische Einheiten. Dies ist ein geringfügiger, aber entscheidender Unterschied. Eine semantische Einheit ist ein Stück Information oder Wissen. Ein Metadatenelement ist eine definierte Vorgabe für die Darstellung dieser Information in einem Metadatensatz, einem Schema oder in einer Datenbank. PREMIS spezifiziert nicht, wie Metadaten in einem System repräsentiert werden sollen, sondern definiert bloß, was das System wissen soll und was es in andere Systeme exportieren können sollte. Um also ein PREMIS-Purist zu sein, müssen Sie in eher abstrakten semantischen Einheiten denken. Die semantischen Einheiten von PREMIS können direkt in die Metadatenelemente, die im PREMIS XML-Schema definiert sind, gemappt werden und möglicherweise auch auf weniger direktem Wege zu Metadaten in anderen Schemata.

Die Namen der semantischen Einheiten von PREMIS sind „camelCase“-Strings. Das heißt, die Begriffe sind nicht durch einen Abstand voneinander getrennt, sondern durch Großbuchstaben: `objectIdentifier` (objektIdentifikator), `relatedEventIdentification` (zugehöriges Ereignis Identifikator). In diesem Dokument sind sie in der Schriftart Verdana gedruckt.

2.2. Container und untergeordnete Einheiten

Einige der semantischen Einheiten sind als *Container* definiert, was bedeutet, dass sie nicht selbst mit einem Wert belegt werden, sondern der Gruppierung von zusammengehörenden semantischen Einheiten dienen. So müssen Sie beispielsweise, wenn Sie einen Identifikator in PREMIS verzeichnen, angeben um was für eine Art von Identifikator es sich handelt (z.B. „DOI“, „ISBN“, „vom lokalen System vergeben“). Der Container `objectIdentifier` (objektIdentifikator) wird verwendet, um die zwei Untereinheiten `objectIdentifierType` (objektIdentifikatorTyp) und `objectIdentifierValue` (objektIdentifikatorWert) zusammenzuführen.

Die Container geben dem Data Dictionary eine hierarchische Struktur, die sich in Version 2.0 in der Nummerierung der semantischen Einheiten widerspiegelt.

1.1 `objectIdentifier` (objektIdentifikator) (M, R)

1.1.1 `objectIdentifierType` (objektIdentifikatorTyp) (M, NR)

1.1.2 `objectIdentifierValue` (objektIdentifikatorWert) (M, NR)

Dieser Auszug aus dem Data Dictionary zeigt auf einen Blick, dass die semantische Einheit `objectIdentifier` verpflichtend (M; mandatory) und wiederholbar (R; repeatable) ist. Da unterhalb

davon weitere semantische Einheiten definiert sind, kann daraus abgeleitet werden, dass *objectIdentifier* nicht selbst mit einem Wert belegt ist, sondern als Container für die zugehörigen Elemente *objectIdentifierType* und *objectIdentifierValue* dient. Da *objectIdentifierType* und *objectIdentifierValue* innerhalb des Containers nicht wiederholbar (NR; non-repeatable) sind, müsste die gesamte Containerstruktur wiederholt werden, falls zwei verschiedene Identifikatoren verzeichnet werden sollen.

2.3. Erweiterungscontainer

Ein *Erweiterungscontainer* ist eine spezielle Form von Container, für den keine Untereinheiten definiert sind. Er ist als Ort dafür gedacht, Metadaten, die nicht aus PREMIS stammen, zu verzeichnen. Auf diese Weise kann PREMIS so erweitert werden, dass auch Metadaten integriert werden können, die nicht im Anwendungsbereich liegen oder aus anderen Gründen nicht in das Data Dictionary aufgenommen wurden. Als letzter Teil ist im Namen des Erweiterungscontainers „Extension“ (Erweiterung) enthalten.

So sind zum Beispiel formatspezifische technische Metadaten nicht in PREMIS enthalten, obwohl sie für die Langzeitarchivierung sehr wichtige Informationen darstellen. Der Erweiterungscontainer *objectCharacteristicsExtension* (objektMerkmaleErweiterung) bietet einen Platz für die Verzeichnung technischer Metadaten, wie sie in anderen Data Dictionaries, wie dem Z39.97-Standard für die Beschreibung von Bitmap Images, definiert sind.

Falls Sie mit XML vertraut sind, wird es für Sie offensichtlich sein, dass das PREMIS Data Dictionary so gestaltet wurde, dass es mit XML kompatibel ist. Die semantischen Einheiten von PREMIS können als XML-Elemente implementiert werden; Containereinheiten sind Elemente, die nur andere Elemente zum Inhalt haben, und Erweiterungseinheiten sind Container für Elemente, die durch externe Schemata definiert sind. Weitere Informationen zu PREMIS und XML finden sich weiter unten in Kapitel 5.1.

3. DAS PREMIS DATENMODELL

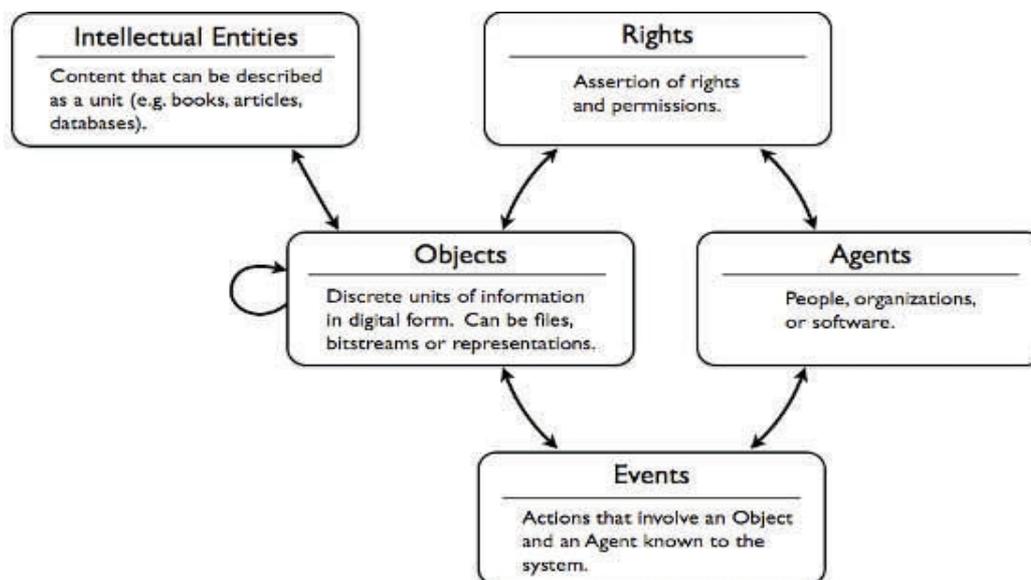


Abbildung 2: Das PREMIS Datenmodell

Eines der Hauptprinzipien hinter PREMIS ist, dass man sich sehr klar darüber sein muss, was man beschreibt. PREMIS definiert fünf Dinge (so genannte *Entitäten* [Entities]), die beschrieben werden können: Intellektuelle Entitäten (Intellectual Entities), Objekte (Objects), Agenten (Agents), Ereignisse

(Events) und Rechte (Rights). Dies wird als das PREMIS-Datenmodell bezeichnet, wie es oben in Abbildung 2 zu sehen ist.

3.1. Intellektuelle Entität

Intellektuelle Entitäten sind konzeptionell und könnten „bibliographische Entitäten“ genannt werden. PREMIS definiert eine Intellektuelle Entität als „ein Set von Inhalten, das zu Zwecken des Managements und der Beschreibung als eine einzige intellektuelle Einheit betrachtet wird: Zum Beispiel ein bestimmtes Buch, eine Landkarte, Fotografie oder eine Datenbank.“ PREMIS definiert tatsächlich aber keine Metadaten, die die Intellektuelle Entität betreffen, da bereits zahlreiche deskriptive Metadatenstandards zur Verfügung stehen.

PREMIS gibt aber vor, dass ein Objekt in einem Archivierungssystem mit der Intellektuellen Entität, die es repräsentiert, verknüpft sein soll, indem es einen Identifikator für die Intellektuelle Entität in die Metadaten des Objekts aufnimmt. Wenn man zum Beispiel eine Kopie des Buches *Buddhism: The EBook: an Online Introduction* langzeitarchiviert, kann man die ISBN als die Verknüpfung der Intellektuellen Entität in der PREMIS Beschreibung des E-Books verwenden.

3.2. Entität Objekt

Objekte sind das, was tatsächlich in einem digitalen Langzeitarchiv gespeichert und verwaltet wird. Der größte Teil von PREMIS widmet sich der Beschreibung von digitalen Objekten. Zu den Informationen, die verzeichnet werden können, gehören:

- ein eindeutiger Identifikator für das Objekt (Typ und Wert),
- Informationen über die Datenintegrität wie z.B. eine Checksumme und der Algorithmus mit dem diese erzeugt wurde,
- die Größe des Objekts,
- das Format des Objekts, das direkt angegeben werden kann oder über einen Link zu einer Format-Registry (Formatverzeichnis),
- der ursprüngliche Name des Objekts,
- Informationen über die Entstehung des Objekts,
- Informationen über Zugriffssperren,
- Informationen über die signifikanten Eigenschaften des Objekts,
- Informationen über die technische Umgebung des Objekts (siehe unten),
- wo und auf welchem Medium das Objekt gespeichert ist,
- Informationen über digitale Signaturen,
- Beziehungen zu anderen Objekten und anderen Formen von Entitäten.

Es sind mehrere semantische Einheiten definiert, um die *technische Umgebung* eines Objekts, d.h. die zur Darstellung bzw. Anwendung benötigte Hard- und Software und die bestehenden Abhängigkeiten zu anderen Objekten, verzeichnen zu können. So kann zum Beispiel eine PDF-Datei sowohl mit unterschiedlichen Versionen des Adobe Acrobat und des Adobe Reader als auch mit anderen Open Source sowie kommerziellen Programmen angezeigt werden. Jedes dieser Programme wird wiederum von unterschiedlichen Betriebssystemen unterstützt und verlangt ein Minimum an Hardware-Voraussetzungen (Prozessorengeschwindigkeit, Arbeitsspeicher und Festplattenkapazität). Da der Adobe Reader keine eigenständige Anwendung ist, sondern ein Browser-Plug-in, ist er wiederum abhängig von bestimmten Versionen bestimmter Browser: Reader 9 für Mac OS benötigt beispielsweise den Safari Browser 2.0.4 oder einer spätere Version davon.

Informationen über die technische Umgebung sind entscheidend für bestimmte Langzeitarchivierungsstrategien. Simple Informationen zur technischen Umgebung, wie z. B. jene für eine einzelne funktionierende Umgebung können in PREMIS einfach verzeichnet werden, wohingegen die Bestimmung von umfassenderen Umgebungsinformationen schwierig und zeitaufwändig sein kann. Falls es detaillierte Informationen gibt, die sich auf ganze Klassen von

Objekten und nicht nur auf lokal vorliegende Objekte beziehen, wie das für Umgebungsinformationen der Fall ist, können die Informationen sinnvollerweise in so genannten Registries (Verzeichnissen) gesammelt und mit allen anderen Langzeitarchiven geteilt werden. PREMIS erlaubt es, zu Informationen, die in externen Registries vorliegen, zu verlinken, wenn dies einer lokalen Verzeichnung vorzuziehen ist.

In PREMIS sind drei unterschiedliche Typen von Objekten definiert und es ist erforderlich, dass die Anwender eine Unterscheidung zwischen diesen treffen. Es gibt die Objekte *Datei (File)*, *Repräsentation (representation)* und *Bitstream (bitstream)*.

Das Datei-Objekt meint genau das, nach dem es klingt: eine Computer-Datei, wie ein PDF oder eine JPEG-Datei.

Ein Objekt vom Typ Repräsentation steht für die Menge aller Objekte vom Typ Datei, die man braucht um eine Intellektuelle Entität darzustellen. Nehmen wir beispielsweise mal an, Sie wollen eine Website archivieren, vielleicht die Homepage ihrer eigenen Institution zu einem bestimmten Zeitpunkt. Die Wahrscheinlichkeit ist hoch, dass die Homepage, wie Sie sie in Ihrem Browser sehen, sich faktisch aus einer Vielzahl von Dateien zusammensetzt – einer oder mehrerer HTML-Dateien, einer Handvoll GIF- oder JPEG-Images, möglicherweise einer Audiodatei oder einer Flash-Animation. Es wird wahrscheinlich auch ein Stylesheet eingesetzt um die Darstellung, die Sie sehen, zu erzeugen. All diese Dateien zusammen werden also benötigt, um die Homepage im Browser zur Ansicht zu bringen. Will ein Archiv die Homepage in einer darstellbaren Form langzeitarchivieren, muss es Kenntnis von all diesen Dateien haben und wissen, wie diese zusammengesetzt sind. Das Repräsentationsobjekt ermöglicht es dem Langzeitarchiv nicht nur die Menge aller zusammengehörenden Dateien zu bestimmen, sondern auch die Merkmale ihres Gesamtzusammenhangs (z. B. also der Website als Ganzes) zu beschreiben, da sich dieser ja von den Merkmalen der einzelnen Teile durchaus unterscheiden kann.

Bitstream-Objekte sind Teilmengen von Dateien. Ein Bitstream-Objekt ist definiert als die Daten (Bits) innerhalb einer Datei, die a) über gemeinsame Eigenschaften hinsichtlich der Archivierung verfügen und b) nicht für sich alleine stehen können, d.h. ohne dass sie durch einen File Header oder eine andere Struktur ergänzt werden. Wenn Sie zum Beispiel eine Datei im AVI-Format (Audio-Video Interleaved) haben, möchten Sie möglicherweise zwischen Audio-Bitstream und Video-Bitstream unterscheiden und diese jeweils als separate Bitstream-Objekte beschreiben.

Einige der im PREMIS Data Dictionary definierten semantischen Einheiten sind auf alle drei Typen von Objekten anwendbar, während andere nur für einen oder zwei Typen von Objekten verwendet werden können. Dadurch, dass es verschiedene Typen von Objekten gibt, ist man gezwungen, sich gründlich darüber Gedanken zu machen, was man beschreibt und dabei so genau wie möglich vorzugehen, da dies für die maschinelle Verarbeitung von Bedeutung ist.

3.3. Ereignisse

Die *Entität Ereignis* führt Informationen über Handlungen zusammen, die die Objekte im digitalen Langzeitarchiv betreffen. Eine genaue und vertrauenswürdige Dokumentation der Ereignisse ist entscheidend, um die digitale Provenienz eines Objekts bestimmen zu können. Diese ist wiederum wichtig, um die Authentizität eines Objekts nachweisen zu können.

Zu den Informationen, die über Ereignisse verzeichnet werden können, gehören:

- ein eindeutiger Identifikator für das Ereignis (Typ und Wert),
- der Typ des Ereignisses (Erstellung, Aufnahme, Migration, etc.),
- Datum und Zeitpunkt an dem das Ereignis stattgefunden hat,
- eine detaillierte Beschreibung des Ereignisses,
- eine codierte Beschreibung des Ergebnisses des Ereignisses,

- eine detaillierte Beschreibung des Ergebnisses des Ereignisses,
- die am Ereignis beteiligten Agenten und ihre Rollen,
- die am Ereignis beteiligten Objekte und ihre Rollen.

Jedes Langzeitarchiv muss für sich selbst entscheiden, welche Ereignisse dauerhaft als Teil der Historie eines Objekts verzeichnet werden sollen. PREMIS empfiehlt, dass Handlungen, die ein Objekt verändern, immer verzeichnet werden sollten. Der Eintrag für `eventType` (`ereignisTyp`) im Data Dictionary enthält des Weiteren eine „Einstiegsliste“ mit wichtigen Typen von Ereignissen, um Langzeitarchive dabei zu unterstützen, diese Ereignisse einheitlich zu verzeichnen.

3.4. Agenten

Agenten sind Handelnde, die bestimmte Rollen bei Ereignissen und in Bezug auf rechtlichen Angaben (siehe 3.5. Rechte) haben. Agenten können Personen, Organisationen oder Softwareanwendungen sein. PREMIS definiert nur eine Minimalzahl an semantischen Einheiten für die Bestimmung von Agenten, da es bereits zahlreiche externe Standards gibt, die verwendet werden können, um genauere Informationen zu verzeichnen (Eine Auswahl davon findet sich unter: "Metadata standards and specifications for describing people and their interests", verfügbar unter: www.ukoln.ac.uk/metadata/resources/people/.) Ein Langzeitarchiv kann sich dafür entscheiden, einen externen Standard für die Verzeichnung von zusätzlichen Informationen über Agenten auszuwählen oder aber es benutzt den Agenten-Identifikator, um auf extern verzeichnete Informationen zu verweisen.

Das Data Dictionary beinhaltet:

- einen eindeutigen Identifikator für den Agenten (Typ und Wert),
- den Namen des Agenten,
- Bestimmung der Art des Agenten (Person, Organisation, Software).

Immer wenn ein Agent in einer Beziehung zu einem Ereignis oder zu rechtlichen Angaben steht, sollte die Rolle des Agenten dabei auch verzeichnet werden. Jeder Agent kann beliebig viele Rollen einnehmen. So kann ich zum Beispiel der Autor und Rechteinhaber eines Werkes sein, der Autor (aber nicht der Rechteinhaber) eines zweiten Werkes und der Ablieferer eines dritten Werkes. Im PREMIS-Modell würde ein Langzeitarchiv einen eindeutigen Identifikator für mich vergeben und diesen Identifikator in allen Ereignis-Einträgen oder rechtlichen Aussagen referenzieren, in denen ich als Agent aufgetreten bin, zusammen mit meiner Rolle im jeweiligen Kontext. Die Rolle des Agenten in Bezug auf ein Ereignis oder rechtliche Aussagen wird dabei als eine Eigenschaft der Entität Ereignis oder der Entität Rechte angesehen, nicht jedoch als eine der Entität Agent selbst.

3.5. Rechte

Die meisten Langzeitarchivierungsstrategien umfassen das Anfertigen von identischen Kopien und abgeleiteter Versionen digitaler Objekte – Handlungen, die laut Urheberrecht den Rechteinhabern vorbehalten sind. Die Entität *Rechte* fasst Informationen über rechtliche Angaben und Erlaubnisse zusammen, die unmittelbar für die Erhaltung von Objekten in einem Langzeitarchiv relevant sind. Alle rechtlichen Angaben in PREMIS halten zwei Dinge fest: Handlungen, für deren Durchführung das Langzeitarchiv über die entsprechenden Rechte verfügt und die Grundlage, aufgrund derer diese Rechte in Anspruch genommen werden.

So könnte ein Langzeitarchiv zum Beispiel eine eingescannte Version eines Buches besitzen, das 1848 veröffentlicht wurde und deshalb frei von Urheberrechten ist. Das Langzeitarchiv kann aufgrund dieses urheberrechtlichen Status mit der digitalen Version des Objekts machen was es will. Ein anderes Langzeitarchiv besitzt ein Objekt, das von einer veröffentlichten CD kopiert wurde, wobei die Lizenz es erlaubt, Back-up-Kopien anzufertigen, aber den Zugang und Gebrauch einschränkt.

Die Informationen, die in den Aussagen über die Rechte verzeichnet werden können, umfassen:

- einen eindeutigen Identifikator für die rechtlichen Angaben (Typ und Wert),
- ob die Grundlage für die Inanspruchnahme des Rechtes das Urheberrecht, eine Lizenzbestimmung oder andere gesetzliche Bestimmungen sind,
- detailliertere Informationen über den urheberrechtlichen Status, Lizenz- oder gesetzliche Bestimmungen, sofern sie Anwendung finden,
- die Handlungen, die aufgrund der rechtlichen Angaben erlaubt sind,
- jegliche Einschränkung dieser Handlungen,
- den Zeitraum der Einräumung oder den Zeitraum für den die rechtlichen Angaben Anwendung finden,
- das bzw. die Objekt(e), für die die rechtlichen Angaben Anwendung finden,
- die an den rechtlichen Angaben beteiligten Agenten und ihre Rollen.

Der Großteil der Informationen ist *ausführbar* (d.h. in einer kontrollierten Form verzeichnet, die eine Verarbeitung durch Computerprogramme ermöglicht) gestaltet. Die rechtlichen Angaben in PREMIS sind Feststellungen der Rechte, nicht ein Verzeichnis von Informationen aufgrund derer die Rechte erst bestimmt werden können. Das bedeutet, dass in PREMIS nicht jene Art von detaillierten Informationen über Autoren, Datum und Ort der Veröffentlichung sowie urheberrechtliche Hinweise definiert sind, wie sie sich in der Copyright-Metadaten-Spezifikation der California Digital Library (www.cdlib.org/inside/projects/rights/schema/) finden. Während es der Sinn von Copyright-Metadaten ist, es Menschen zu ermöglichen, fortlaufend den rechtlichen Status bestimmen zu können, ist es der Sinn der Entität Rechte in PREMIS, ausführbare Informationen für Langzeitarchivierungssysteme zur Verfügung zu stellen.

4. DAS DATA DICTIONARY

4.1. Beispieleintrag im Data Dictionary für eine einfache semantische Einheit

Abbildung 3 zeigt den Eintrag des Data Dictionary für die semantische Einheit *size* (Datenmenge), die eine Komponente oder Untereinheit des Containers mit der Bezeichnung *objectCharacteristics* (objektMerkmale) ist. *size* selbst hat keine Untereinheiten. Der Eintrag des Data Dictionary beinhaltet eine Definition des Elements und einen Grund (Begründung) für die Aufnahme ins Kernset der PREMIS Metadaten, sowie Beispiele und Notizen, wie dieser Wert belegt und genutzt werden kann. Dies soll den Anwendern dabei helfen, das Element richtig zu verwenden.

Die zwei Zeilen „Objekt-Kategorie“ (Object Category) und „Anwendbarkeit“ (Applicability) werden gemeinsam benötigt, um darzustellen, ob die semantische Einheit geeignet ist, Repräsentationen, Dateien und/ oder Bitstreams zu beschreiben. Hier bezieht sich *size* lediglich auf Dateien und Bitstreams. Schließlich gibt es ein Set von Regeln für die Anwendung: „Datenbeschränkungen“ (Data constraint), „Wiederholbarkeit“ (Repeatability) und „Verpflichtung“ (Obligation).

Datenbeschränkungen spezifizieren Einschränkungen hinsichtlich der Werte, die eine semantische Einheit annehmen kann. In unserem Beispiel muss der Wert von *size* eine ganze Zahl sein. Eine andere übliche Einschränkung der Daten ist, dass der Wert aus einem kontrollierten Vokabular stammen muss. Manchmal sind die Begriffe des Vokabulars im Data Dictionary spezifiziert und manchmal nicht, aber auf alle Fällen muss der Name des Vokabulars verzeichnet werden. Im Data Dictionary sind keine semantischen Einheiten für die Namen von Vokabularen definiert, aber die PREMIS Maintenance Activity ist dabei, eine Lösung für das XML-Schema zu entwickeln.

Wiederholbarkeit gibt an, ob eine semantische Einheit wiederholbar ist.

Verpflichtung gibt an, ob ein Wert für eine semantische Einheit verpflichtend (erforderlich) oder optional ist. Verpflichtung ist möglicherweise verwirrend, da PREMIS eindeutig aussagt, dass es von Langzeitarchiven nicht fordert, dass sie irgendwelche speziellen Informationen verzeichnen und

speichern müssen. Eine verpflichtende semantische Einheit muss nicht verzeichnet und im Langzeitarchiv abgelegt sein. Dennoch muss das Langzeitarchiv in der Lage sein, den Wert der semantischen Einheiten zu erzeugen, wenn er benötigt wird, zum Beispiel für den Austausch mit einem anderem Langzeitarchiv. Für den unwahrscheinlichen Fall, das z. B. ein Langzeitarchiv nichts anderes außer TIFF 6.0 Images speichert, müsste es nicht für jedes Objekt Formatinformationen verzeichnen. Nichtsdestotrotz würde das Management des Langzeitarchivs wissen, dass es sich bei den Dateien um TIFF 6.0 Images handelt und könnte diese Information weitergeben, wenn es müsste. (Siehe Kapitel 5.2 PREMIS Konformität).

Semantic Unit (Semantische Einheit)	1.5.3. size (1.5.3. datenMenge)		
Semantic Components (Sematische Komponenten)	None (Keine)		
Definition (Definition)	Die Größe der im Langzeitarchiv gespeicherten Datei oder des Bitstreams in Bytes		
Rationale (Begründung)	datenMenge stellt sicher, dass aus dem Langzeitarchiv die korrekte Anzahl an Bytes abgerufen wird und dass eine Anwendung über genügend Speicherplatz verfügt, um Dateien zu verschieben oder zu verarbeiten. Es kann auch zur Berechnung von Speicherkosten verwendet werden.		
Data constraint (Datenbeschränkung)	Integer (Ganze Zahlen)		
Object Category (Objektkategorie)	Repräsentation	Datei	Bitstream
Applicability (Anwendbarkeit)	Not applicable (Nicht anwendbar)	Applicable (Anwendbar)	Applicable (Anwendbar)
Examples (Beispiele)		2038937	
Repeatability (Wiederholbarkeit)		Not repeatable (Nicht wiederholbar)	Not repeatable (Nicht wiederholbar)
Obligation (Verpflichtung)		Optional (optional)	Optional (optional)
Creation / Maintenance Notes (Bemerkungen zur Entstehung / Erhaltung und Pflege)	Automatisch vom Langzeitarchivierungssystem aufgezeichnet.		
Usage Notes (Bemerkungen zum Gebrauch)	Dadurch dass diese semantische Einheit als Größe in Bytes definiert ist, ist es nicht nötig, eine Maßeinheit festzuhalten. Allerdings muss beim Datenaustausch die Maßeinheit benannt und von beiden Partnern verstanden werden.		

Abbildung 3: Auszug aus dem Data Dictionary für die semantische Einheit size

4.2. Beispieleintrag im Data Dictionary für eine Container-Einheit

Abbildung 4 zeigt den Anfang eines Eintrages im Data Dictionary für objectCharacteristics (objectCharakteristika), der Container-Einheit für size. Dass es sich um einen Container handelt,

sieht man daran, dass die Einheit semantische Komponenten hat und die Datenbeschränkung „Container“ ist. Zu beachten ist, dass die enthaltenen semantischen Komponenten unitär sein können, wie *size*, oder selbst wiederum Container sein können, wie *Format*.

Semantic Unit (Semantische Einheit)	1.5. objectCharacteristics (1.5 objektMerkmale)		
Semantic Components (Sematische Komponenten)	1.5.1 compositionLevel (1.5.1 kodierungsLevel) 1.5.2 fixity (1.5.2 datenIntegritaet) 1.5.3 size (1.5.3 Datenmenge) 1.5.4 format (1.5.4 format) 1.5.5 creatingApplication (1.5.5 erzeugendesProgramm) 1.5.6 inhibitors (1.5.6. Zugriffssperren) 1.5.7 objectCharacteristicsExtension (1.5.7. objektMerkmaleErweiterung)		
Definition (Definition)	Technische Eigenschaften einer Datei oder eines Bitstreams, die auf alle oder die meisten Formate zutreffen.		
Rationale (Begründung)	Es gibt einige technische Eigenschaften, die auf Objekte aller Formate anwendbar sind. Eine detaillierte Definition von formatspezifischen Eigenschaften ist nicht Teil des Data Dictionary, solche Eigenschaften können jedoch unter objectCharacteristicsExtension (objektMerkmaleErweiterung) aufgenommen werden.		
Data constraint (Datenbeschränkung)	Container (Container)		
Object Category (Objektkategorie)	Repräsentation	Datei	Bitstream
Applicability (Anwendbarkeit)	Not applicable (Nicht anwendbar)	Applicable (Anwendbar)	Applicable (Anwendbar)
Repeatability (Wiederholbarkeit)		Repeatable (Wiederholbar)	Repeatable (Wiederholbar)
Obligation (Verpflichtung)		Mandatory (Verpflichtend)	Mandatory (Verpflichtend)
Usage Notes (Anmerkungen zum Gebrauch)	Die semantischen Einheiten, die in objectCharacteristics enthalten sind, sollten als ein Set von Informationen verwendet werden, das ein auf einzelnes Objekt auf einem einzigen compositionLevel betrifft. ObjectCharacteristics kann wiederholt werden [...]		

Abbildung 4: Auszug aus dem Data Dictionary für die semantische Einheit objectCharacteristics

5. DIE ANWENDUNG VON PREMIS

5.1. PREMIS in XML

Es wird angenommen (wenngleich nicht zwingend vorausgesetzt), dass PREMIS, sofern es zum Austausch eingesetzt wird, in XML dargestellt wird. Die PREMIS Maintenance Activity bietet ein XML-Schema an, das unmittelbar dem Data Dictionary entspricht, um eine einfache Beschreibung von Objekten, Ereignissen, Agenten und Rechten zu ermöglichen. Abbildung 5 zeigt einen Auszug aus den PREMIS Metadaten unter Einsatz des PREMIS XML-Schemas.

In der Praxis setzen die meisten Archivierungssysteme bereits XML-Formate für den Import und Export von Daten ein. Viele benutzen METS (Metadata Encoding and Transmission Standard) – einen weiteren von der Library of Congress betreuten Standard - als einen XML-Container um unterschiedliche Typen von Metadaten zusammenzuführen. Es ist möglich, PREMIS innerhalb von METS zu benutzen, aber aus zwei Gründen ist es nicht ganz einfach. Erstens, splittet METS Informationen in verschiedene Sektionen auf, je nachdem, ob es sich um technische Metadaten, rechtliche Metadaten oder Provenienz-Metadaten handelt. Das PREMIS Schema, das dem Data Dictionary folgt, hat Abschnitte für Objekte, Rechte, Ereignisse und Agenten. Es gibt einige Entsprechungen zwischen den beiden Strukturen, aber es ist nicht perfekt, insbesondere bezüglich der Informationen über Agenten. Zweitens, haben PREMIS und METS gewisse Überschneidungen, zum Beispiel definieren beide einen Tag für die Verzeichnung von Checksummen. Wenn PREMIS und METS gemeinsam eingesetzt werden, muss man sich entscheiden, ob man diese sich überschneidenden Elemente im PREMIS-Teil, im METS-Teil oder in beiden verzeichnen möchte.

Klar ist, dass wenn jedes Langzeitarchiv diesbezüglich seine eigenen Entscheidungen trifft, es zu einer großen Bandbreite in der Repräsentation der Daten kommen kann, die die Interoperabilität beeinträchtigt. Deshalb sind derzeit eine Reihe von Aktivitäten im Gange, die versuchen, best-practice-Beispiele für einen gemeinsamen Einsatz von PREMIS und METS zu definieren. Über den Fortschritt dieser Bemühungen gibt die Webseite der PREMIS Maintenance Activity Auskunft.

```
<event>
  <eventIdentifier>
    <eventIdentifierType>DAITSS</eventIdentifierType>
    <eventIdentifierValue>10012</eventIdentifierValue>
  </eventIdentifier>
  <eventType>Validation</eventType>
  <eventDateTime>2008-05-06T10:40:22-04:00</eventDateTime>
  <eventOutcomeInformation>
    <eventOutcome>Invalid</eventOutcome>
    <eventOutcomeDetail>ill-formed DateTime
      value</eventOutcomeDetail>
  </eventOutcomeInformation>
</event>
```

Abbildung 5: Ein PREMIS-Auszug in XML

5.2. PREMIS Konformität

Das PREMIS Data Dictionary enthält einen Abschnitt dazu, was es für ein Langzeitarchiv bedeutet, PREMIS-konform zu sein. Im Wesentlichen ergeben sich drei Anforderungen:

1. Wenn das Langzeitarchiv ein Datenelement anwendet (speichert oder exportiert), welches vorgibt, eine semantische Einheit von PREMIS zu sein, sollte dieses Datenelement der gleichen Definition,

den gleichen Datenbeschränkungen und der gleichen Anwendbarkeit unterliegen wie die in PREMIS definierte semantische Einheit.

2. Wenn ein Langzeitarchiv eine semantische Einheit aus PREMIS verwendet, können die Wiederholbarkeit und die Verpflichtung strikter als in PREMIS gehandhabt werden, jedoch nicht freier. Das bedeutet, dass eine wiederholbare semantische Einheit als nicht-wiederholbar eingesetzt werden kann, aber nicht andersherum und dass ein verpflichtendes Element nicht als optional definiert werden kann.

3. Wenn das Langzeitarchiv Informationen für den Gebrauch in einem anderen Langzeitarchiv exportiert, muss es Werte für alle semantischen Einheiten, die im Data Dictionary verpflichtend sind, bereitstellen. Dennoch ist dabei ein wenig Flexibilität gegeben, denn es wird von den Langzeitarchiven nicht gefordert, verpflichtende semantische Einheiten für Typen von Entitäten zu unterstützen, wenn es diese Entitäten selbst nicht unterstützt. In anderen Worten, ein Langzeitarchiv kann selbst entscheiden, ob es PREMIS Agenten unterstützt oder nicht, aber wenn es dies tut, dann ist der agentIdentifier (agentIdentifikator) verpflichtend. In gleicher Weise muss auch ein spezielles Langzeitarchiv, das Bitstream-Objekte nicht unterstützt, für diesen Fall auch nicht den ansonsten verpflichtenden Bitstream-Identifikator angeben.

Diese Anforderungen sollen im Kontext einer Reihe von Dingen betrachtet werden, die für die Konformität NICHT erforderlich sind. Wie oben beschrieben, ist es für ein Langzeitarchiv nicht erforderlich, alle Typen von Entitäten, die im PREMIS Datenmodell definiert sind zu unterstützen. Es ist ebenfalls nicht erforderlich, die Metadaten intern unter dem Namen der semantischen Einheiten von PREMIS zu speichern oder dabei Werte zu verwenden, die den PREMIS Datenbeschränkungen entsprechen. In anderen Worten: Es kommt es also nicht darauf an, wie ein Langzeitarchiv zu einem PREMIS-Wert kommt – indem es ihn unter dem gleichen oder einem anderen Namen speichert, durch das Mapping von einem anderen Wert, durch den Verweis auf eine Registry, durch Schlussfolgerungen, durch einen default-Wert oder auf einem anderem Wege. Solange also ein Langzeitarchiv einen guten PREMIS-Wert liefern kann, wenn dieser benötigt wird, ist es konform.

In der Zukunft wird die PREMIS Maintenance Activity möglicherweise die Konformitätsanforderungen neu betrachten und sie stringenter gestalten, aber mit Version 2.0 ist es ziemlich einfach für ein Langzeitarchiv PREMIS-konform zu sein. Andererseits: Je mehr semantische Einheiten ein Langzeitarchiv unterstützt, um so mehr bringt die Verwendung von PREMIS. Das PREMIS Data Dictionary wurde entwickelt, um den „Kern“ an Informationen zu identifizieren, die ein Langzeitarchiv benötigen wird, um digitale Inhalte dauerhaft zu erhalten. Ein verantwortlich handelndes Langzeitarchiv sollte sich PREMIS genau anschauen und gute Gründe haben, wenn es Teile des Data Dictionary nicht anwendet.

6. WEITERE INFORMATIONEN

Auf der Website der PREMIS Maintenance Activity (www.loc.gov/standards/premis) findet sich für jeden etwas, unter anderem die Links zur PREMIS Anwender Gruppe (PREMIS Implementers Group – PIG), PREMIS-Tutorials, Schemata, Werkzeuge und Neuigkeiten. Es gibt dort auch einen Abschnitt: „Ressourcen: Artikel und Präsentationen“ (Resources: Articles and Presentations), der zu Literatur über PREMIS und verwandten Themen verlinkt. Einige hilfreiche und allgemein verständliche Ressourcen sind hier aufgelistet:

Zu Langzeitarchivierungsmetadaten allgemein:

“Preservation Metadata” (PDF:209K/21pp.)

Brian Lavoie (OCLC) and Richard Gartner (Oxford)

Joint report of OCLC, Oxford Library Services, and the Digital Preservation Coalition. Published as DPC Technology Watch Report No. 05-01: September 2005. <http://www.dpconline.org/docs/reports/dpctw05-01.pdf>

Zu den Unterschieden zwischen den PREMIS Versionen 1.0 und 2.0:

"PREMIS with a Fresh Coat of Paint: Highlights from the Revision of the PREMIS Data Dictionary for Preservation Metadata "

Brian Lavoie, *D-Lib Magazine*, May/June 2008.

<http://www.dlib.org/dlib/may08/lavoie/05lavoie.html>

Zur gemeinsamen Anwendung von PREMIS und METS:

"Battle of the Buzzwords: Flexibility vs. Interoperability When Implementing PREMIS with METS"

Rebecca Guenther, *D-Lib Magazine*, July/August 2008.

<http://www.dlib.org/dlib/july08/guenther/07guenther.html>

Appendix A: Liste aller semantischen Einheiten in PREMIS

Aus dem PREMIS Data Dictionary für Langzeitarchivierungsmetadaten Version 2.0.

Wie diese Liste zu lesen ist

1.2 objectCategory (M, NR)

Die semantische Einheit objectCategory (objektKategorie) ist Verpflichtend (Mandatory, M) und Nicht Wiederholbar (Not Repeatable, NR). Sie findet Anwendung für alle Typen von Objekten (Datei, Repräsentation, Bitstream).

1.3. preservationLevel (O, R) [representation, file]

1.3.1. preservationLevelValue (M, NR) [representation, file]

Die semantische Einheit preservationLevel (archivierungsLevel) ist Optional (Optional, O) und Wiederholbar (Repeatable, R). Sie kann nur für die Objekttypen Repräsentation und Datei eingesetzt werden. Sie ist eine Container-Einheit, da sie mindestens eine untergeordnete semantische Einheit – preservationLevelValue (archivierungsLevelWert), ebenfalls nur auf Repräsentation (representation) und Datei (file) anwendbar – enthält.

Semantische Einheiten der Entität Objekt

1.1 objectIdentifier (objektIdentifikator) (M, R)

1.1.1 objectIdentifierType (objektIdentifikatorTyp) (M, NR)

1.1.2 objectIdentifierValue (objektIdentifikatorWert) (M, NR)

1.2 objectCategory (objektKategorie) (M, NR)

1.3 preservationLevel (archivierungsLevel) (O, R) [representation, file]

1.3.1 preservationLevelValue (archivierungsLevelWert) (M, NR) [representation, file]

1.3.2 preservationLevelRole (archivierungsLevelRolle) (O, NR) [representation, file]

1.3.3 preservationLevelRationale (archivierungsLevelBegrueendung)(O, R) [representation, file]

1.3.4 preservationLevelDateAssigned (archivierungsLevelZugewiesenesDatum) (O, NR)

[representation, file]

1.4 significantProperties (signifikanteEigenschaften) (O, R)

1.4.1 significantPropertiesType (signifikanteEigenschaftenTyp) (O, NR)

1.4.2 significantPropertiesValue (signifikanteEigenschaftenWert) (O, NR)

1.4.3 significantPropertiesExtension (signifikanteEigenschaftenErweiterung)(O, R)

1.5 objectCharacteristics (objektMerkmale) (M, R) [file, bitstream]

1.5.1 compositionLevel (kodierungsLevel) (M, NR) [file, bitstream]

1.5.2 fixity (O, R) (datenIntegritaet) [file, bitstream]

1.5.2.1 messageDigestAlgorithm (pruefsummeAlgorithmus) (M, NR) [file, bitstream]

1.5.2.2 messageDigest (M, NR) (pruefsummeWert) [file, bitstream]

1.5.2.3 messageDigestOriginator (pruefsummeErzeuger) (O, NR) [file, bitstream]

1.5.3 size (datenMenge) (O, NR) [file, bitstream]

1.5.4 format (format) (M, R) [file, bitstream]

- 1.5.4.1 formatDesignation (formatBezeichnung) (O, NR) [file, bitstream]
 - 1.5.4.1.1 formatName (formatName) (M, NR) [file, bitstream]
 - 1.5.4.1.2 formatVersion (formatVersion) (O, NR) [file, bitstream]
- 1.5.4.2 formatRegistry (formatVerzeichnis) (O, NR) [file, bitstream]
 - 1.5.4.2.1 formatRegistryName (formatVerzeichnisName) (M, NR) [file, bitstream]
 - 1.5.4.2.2 formatRegistryKey (formatVerzeichnisIdentifikator) (M, NR) [file, bitstream]
 - 1.5.4.2.3 formatRegistryRole (formatVerzeichnisRolle) (O, NR) [file, bitstream]
- 1.5.4.3 formatNote (formatAnmerkung) (O, R) [file, bitstream]
- 1.5.5 creatingApplication (erzeugendesProgramm) (O, R) [file, bitstream]
 - 1.5.5.1 creatingApplicationName (erzeugendesProgrammName) (O, NR) [file, bitstream]
 - 1.5.5.2 creatingApplicationVersion (erzeugendesProgrammVersion) (O, NR) [file, bitstream]
 - 1.5.5.3 dateCreatedByApplication (erzeugtVonProgrammAm) (O, NR) [file, bitstream]
 - 1.5.5.4 creatingApplicationExtension (erzeugendesProgrammErweiterung) (O, R) [file, bitstream]
- 1.5.6 inhibitors (zugriffssperren) (O, R) [file, bitstream]
 - 1.5.6.1 inhibitorType (zugriffssperrenTyp) (M, NR) [file, bitstream]
 - 1.5.6.2 inhibitorTarget (zugriffssperrenZiel) (O, R) [file, bitstream]
 - 1.5.6.3 inhibitorKey (zugriffssperrenSchluessel) (O, NR) [file, bitstream]
- 1.5.7 objectCharacteristicsExtension (objektMerkmaleErweiterung) (O, R) [file, bitstream]
- 1.6 originalName (originalName) (O, NR) [representation, file]
- 1.7 storage (speicherung) (M, R) [file, bitstream]
 - 1.7.1 contentLocation (speicherOrt) (O, NR) [file, bitstream]
 - 1.7.1.1 contentLocationType (speicherOrtTyp) (M, NR) [file, bitstream]
 - 1.7.1.2 contentLocationValue (speicherOrtWert) (M, NR) [file, bitstream]
 - 1.7.2 storageMedium (speicherMedium) (O, NR) [file, bitstream]
- 1.8 environment (technischeUmgebung) (O, R)
 - 1.8.1 environmentCharacteristic (technischeUmgebungMerkmal) (O, NR)
 - 1.8.2 environmentPurpose (technischeUmgebungZweck) (O, R)
 - 1.8.3 environmentNote (technischeUmgebungAnmerkung)(O, R)
 - 1.8.4 dependency (abhaengigkeit) (O, R)
 - 1.8.4.1 dependencyName (abhaengigkeitName) (O, R)
 - 1.8.4.2 dependencyIdentifier (abhaengigkeitIdentifikator) (O, R)
 - 1.8.4.2.1 dependencyIdentifierType (abhaengigkeitIdentifikatorTyp) (M, NR)
 - 1.8.4.2.2 dependencyIdentifierValue (abhaengigkeitIdentifikatorWert) (M, NR)
 - 1.8.5 software (software) (O, R)
 - 1.8.5.1 swName (swName) (M, NR)
 - 1.8.5.2 swVersion (swVersion) (O, NR)
 - 1.8.5.3 swType (swTyp) (M, NR)
 - 1.8.5.4 swOtherInformation (swWeitereInformation) (O, R)

- 1.8.5.5 swDependency (swAbhaengigkeit) (O, R)
- 1.8.6 hardware (hardware) (O, R)
 - 1.8.6.1 hwName (hwName) (M, NR)
 - 1.8.6.2 hwType (hwTyp) (M, NR)
 - 1.8.6.3 hwOtherInformation (hwWeitereInformation) (O, R)
- 1.8.7 environmentExtension (technischeUmgebungErweiterung) (O, R)
- 1.9 signatureInformation (signaturInformation) (O, R) [file, bitstream]
 - 1.9.1 signature (signatur) (O, R)
 - 1.9.1.1 signatureEncoding (signaturKodierung) (M, NR) [file, bitstream]
 - 1.9.1.2 signer (O, NR) (signierender) [file, bitstream]
 - 1.9.1.3 signatureMethod (signaturMethode) (M, NR) [file, bitstream]
 - 1.9.1.4 signatureValue (signaturWert) (M, NR) [file, bitstream]
 - 1.9.1.5 signatureValidationRules (signaturValidierungsRegeln) (M, NR) [file, bitstream]
 - 1.9.1.6 signatureProperties (O, R) (signaturEigenschaften) [file, bitstream]
 - 1.9.1.7 keyInformation (O, NR) (schluessellInformation) [file, bitstream]
 - 1.9.2 signatureInformationExtension (signaturInformationErweiterung) (O, R) [file, bitstream]
- 1.10 relationship (objektBeziehungen) (O, R)
 - 1.10.1 relationshipType (objektBeziehungenTyp) (M, NR)
 - 1.10.2 relationshipSubType (objektBeziehungenUnterTyp) (M, NR)
 - 1.10.3 relatedObjectIdentification (beziehungsObjektIdentifikator) (M, R)
 - 1.10.3.1 relatedObjectIdentifierType (beziehungsObjektIdentifikatorTyp) (M, NR)
 - 1.10.3.2 relatedObjectIdentifierValue (beziehungsObjektIdentifikatorWert) (M, NR)
 - 1.10.3.3 relatedObjectSequence (beziehungsObjektSequenz) (O, NR)
 - 1.10.4 relatedEventIdentification (beziehungsEreignisIdentifikator) (O, R)
 - 1.10.4.1 relatedEventIdentifierType (beziehungsEreignisIdentifikatorTyp) (M, NR)
 - 1.10.4.2 relatedEventIdentifierValue (beziehungsEreignisIdentifikatorWert) (M, NR)
 - 1.10.4.3 relatedEventSequence (beziehungsEreignisSequenz) (O, NR)
- 1.11 linkingEventIdentifier (verknuepftesEreignisIdentifikator)(O, R)
 - 1.11.1 linkingEventIdentifierType (verknuepftesEreignisIdentifikatorTyp) (M, NR)
 - 1.11.2 linkingEventIdentifierValue (verknuepftesEreignisIdentifikatorWert) (M, NR)
- 1.12 linkingIntellectualEntityIdentifier (verknuepftelIntellektuelleEntitaetIdentifikator) (O, R)
 - 1.12.1 linkingIntellectualEntityIdentifierType (verknuepftelIntellektuelleEntitaetIdentifikatorTyp) (M, NR)
 - 1.12.2 linkingIntellectualEntityIdentifierValue (verknuepftelIntellektuelleEntitaetIdentifikatorWert) (M, NR)
- 1.13 linkingRightsStatementIdentifier (verknuepfterechtlicheAngabenIdentifikator) (O, R)
 - 1.13.1 linkingRightsStatementIdentifierType (verknuepfterechtlicheAngabenIdentifikatorTyp) (M, NR)
 - 1.13.2 linkingRightsStatementIdentifierValue (verknuepfterechtlicheAngabenIdentifikatorWert) (M, NR)

Semantische Einheiten der Entität Ereignis

- 2.1 eventIdentifier (ereignisIdentifikator) (M, NR)
 - 2.1.1 eventIdentifierType (ereignisIdentifikatorTyp) (M, NR)
 - 2.1.2 eventIdentifierValue (ereignisIdentifikatorWert) (M, NR)
- 2.2 eventType (ereignisTyp) (M, NR)
- 2.3 eventDateTime (ereignisDatumZeit) (M, NR)
- 2.4 eventDetail (ereignisDetail) (O, NR)
- 2.5 eventOutcomeInformation (ereignisResultatInformation) ((O, R))
 - 2.5.1 eventOutcome (ereignisResultat) (O, NR)
 - 2.5.2 eventOutcomeDetail (ereignisResultatDetail) (O, R)
 - 2.5.2.1 eventOutcomeDetailNote (ereignisResultatDetailAnmerkung) (O, NR)
 - 2.5.2.2 eventOutcomeDetailExtension (ereignisResultatDetailErweiterung) (O, R)
- 2.6 linkingAgentIdentifier (verknuepfterAgentIdentifikator) (O, R)
 - 2.6.1 linkingAgentIdentifierType (verknuepfterAgentIdentifikatorTyp) (M, NR)
 - 2.6.2 linkingAgentIdentifierValue (verknuepfterAgentIdentifikatorWert) (M, NR)
 - 2.6.3 linkingAgentRole (verknuepfterAgentIdentifikatorRolle) (O, R)
- 2.7 linkingObjectIdentifier (verknuepftesObjektIdentifikator) (O, R)
 - 2.7.1 linkingObjectIdentifierType (verknuepftesObjektIdentifikatorTyp) (M, NR)
 - 2.7.2 linkingObjectIdentifierValue (verknuepftesObjektIdentifikatorWert) (M, NR)
 - 2.7.3 linkingObjectRole (verknuepftesObjektRolle) (O, R)

Semantische Einheiten der Entität Agent

- 3.1 agentIdentifier (agentIdentifikator) (R, M)
 - 3.1.1 agentIdentifierType (agentIdentifikatorTyp) (M, NR)
 - 3.1.2 agentIdentifierValue (agentIdentifikatorWert) (M, NR)
- 3.2 agentName (agentName) (O, R)
- 3.3 agentType (agentTyp) (O, NR)

Semantische Einheiten der Entität Rechte

- 4.1 rightsStatement (rechtlicheAngaben) (O, R)
 - 4.1.1 rightsStatementIdentifier (rechtlicheAngabenIdentifikator) (M, NR)
 - 4.1.1.1 rightsStatementIdentifierType (rechtlicheAngabenIdentifikatorTyp) (M, NR)
 - 4.1.1.2 rightsStatementIdentifierValue (rechtlicheAngabenIdentifikatorWert) (M, NR)
 - 4.1.2 rightsBasis (rechteGrundlagen) (M, NR)
 - 4.1.3 copyrightInformation (urheberrechtlicheInformation) (O, NR)
 - 4.1.3.1 copyrightStatus (urheberrechtlicherStatus) (M, NR)
 - 4.1.3.2 copyrightJurisdiction (urheberrechtlicheZustaendigkeit) (M, NR)
 - 4.1.3.3 copyrightStatusDeterminationDate (urheberrechtlicherStatusBestimmungsDatum) (O, NR)
 - 4.1.3.4 copyrightNote (urheberrechtAnmerkung) (O, R)

- 4.1.4 licenseInformation (lizenzInformation) (O, NR)
 - 4.1.4.1 licenseIdentifier (lizenzIdentifikator) (O, NR)
 - 4.1.4.1.1 licenseIdentifierType (lizenzIdentifikatorTyp) (M, NR)
 - 4.1.4.1.2 licenseIdentifierValue (lizenzIdentifikatorWert) (M, NR)
 - 4.1.4.2 licenseTerms (lizenzBedingungen) (O, NR)
 - 4.1.4.3 licenseNote (lizenzAnmerkung) (O, R)
- 4.1.5 statuteInformation (gesetzlicheBestimmungenInformation) (O, R)
 - 4.1.5.1 statuteJurisdiction (gesetzlicheBestimmungenZustaendigkeit) (M, NR)
 - 4.1.5.2 statuteCitation (gesetzlicheBestimmungenLiteraturstelle) (M, NR)
 - 4.1.5.3 statuteInformationDeterminationDate (gesetzlicheBestimmungenBestimmungsDatum) (O, NR)
 - 4.1.5.4 statuteNote (gesetzlicheBestimmungenAnmerkungen) (O, R)
- 4.1.6 rightsGranted (rechteEinraeumung) (O, R)
 - 4.1.6.1 act (maßnahme) (M, NR)
 - 4.1.6.2 restriction (einschraenkung) (O, R)
 - 4.1.6.3 termOfGrant (zeitraumDerEinraeumung) (M, NR)
 - 4.1.6.3.1 startDate (startDatum) (M, NR)
 - 4.1.6.3.2 endDate (endDatum) (O, NR)
 - 4.1.6.4 rightsGrantedNote (rechteEinraeumungAnmerkung) (O, R)
- 4.1.7 linkingObjectIdentifier (verknuepftesObjektIdentifikator) (O, R)
 - 4.1.7.1 linkingObjectIdentifierType (verknuepftesObjektIdentifikatorTyp) (M, NR)
 - 4.1.7.2 linkingObjectIdentifierValue (verknuepftesObjektIdentifikatorWert) (M, NR)
- 4.1.8 linkingAgentIdentifier (verknuepfterAgentIdentifikator) (O, R)
 - 4.1.8.1 linkingAgentIdentifierType (verknuepfterAgentIdentifikatorTyp) (M, NR)
 - 4.1.8.2 linkingAgentIdentifierValue (verknuepfterAgentIdentifikatorWert) (M, NR)
 - 4.1.8.3 linkingAgentRole (verknuepfterAgentIdentifikatorRolle) (M, NR)
- 4.2 rightsExtension (rechteErweiterung) (O, R)

Appendix B: Beispiel für ein Objekt

Dies ist die vereinfachte Version eines Beispiels, das die Library of Congress (LoC) für ein PREMIS-Tutorial vorbereitet hat. Es zeigt die semantischen Einheiten und Werte von PREMIS, die zur Beschreibung eines TIFF-Images benutzt werden.

Die LoC setzt *handles* (Identifikatoren, die von einer Anwendung namens *Handle System* erstellt und verwaltet werden) für Datei-Objekte, die im Langzeitarchiv gespeichert werden, ein (siehe 1.1. objectIdentifier und 1.10.3 relatedObjectIdentification). Das beschriebene Objekt ist innerhalb des Langzeitarchivs über seinen eindeutigen Identifikator (1.1 objectIdentifier) auffindbar, aber der ursprüngliche Name vor dem Übergang in das Archiv war „001h.tif“ (1.6 originalName). Aufgrund der Objektkategorie (1.2 objectCategory) wissen wir, dass es sich um eine Datei handelt.

Die Datei hat eine MD5-Checksumme, die vom Digital Content Management System der LoC berechnet wurde (1.5.2 fixity). Die Formatinformation sind auf zwei Arten angegeben, als MIME-Typ und Version (1.5.4.1 formatDesignation) und als Verweis auf die PRONOM-Formatregistry, die eine Kopie der TIFF-Spezifikation beinhaltet.

Die LoC geht von zwei für die Erstellung eingesetzten Anwendungen aus. Etwas (wahrscheinlich eine namenlose RAW-Datei) wurde auf einem Scanner mit der Anwendung ScandAll erzeugt und dann mit Adobe Photoshop in ein TIFF umgewandelt (1.5.5 creatingApplication). Es gibt keine Verschlüsselung oder sonstigen Arten von Zugriffssperren, die aus Sicht der LoC explizit zu verzeichnen gewesen wären (1.5.6 inhibitors).

Die Datei ist auf einem Ort auf Festplatte gespeichert, der mittels dem Verzeichnispfad identifizierbar ist (1.7.1. contentLocation); das Festplattensystem ist sehr genau identifiziert, so dass die Manager des Langzeitarchivs voraussichtlich wissen, wann die Daten auf ein aktuelleres System zu migrieren sind.

Die LoC hat einige Sorgfalt darauf verwendet, eine empfohlene technische Umgebung für das Rendering bzw. die Anzeige der Datei zu verzeichnen: Einsatz von Adobe Acrobat Version 5.0 auf einem Intel x86 Computer unter Windows XP. Die Geschwindigkeit des Prozessors muss mindestens 60 MHz sein und es sollten mindestens 32 MB an Speicher vorhanden sein, wobei 64 MB empfohlen sind (1.8. environment).

Die TIFF-Datei ist Teil einer Repräsentation mit dem intern vergebenen Identifikator R200802948. Diese Information findet sich im dritten Block der Informationen zu Beziehungen (1.10 relationship). Es gibt zwei weitere Dateien in dieser Repräsentation, die mit *handles* identifiziert werden. Wir wissen, dass dies Objekte vom Typ Datei sein müssen, da ihre Form von Beziehungen als „Geschwister“ (sibling) beschrieben sind. Für weitere Informationen, wie um was für Typen von Dateien es sich handelt, müssten wir in der Objektbeschreibung *loc.music/gottlieb.09602* und *loc.music/gottlieb.mets* nachsehen.

Wir haben keine beschreibenden Informationen über die intellektuelle Entität, die durch diese Repräsentation repräsentiert wird. Allerdings haben wir einen Link zu ihrem Identifikator, einer LCCN (1.12 linkingIntellectualEntityIdentifier). Wahrscheinlich würden wir, wenn wir diese LCCN nachschlagen würden, eine bibliographische Beschreibung der intellektuellen Entität finden (in diesem Fall allerdings nicht, da die LCCN frei erfunden ist).

Semantische Einheit	Wert
1.1 objectIdentifier (objektIdentifikator)	
1.1.1 objectIdentifierType (objektIdentifikatorTyp)	hdl
1.1.2 objectIdentifierValue (objektIdentifikatorWert)	loc.music/gottlieb.09601
1.2 objectCategory (objektKategorie)	file
1.3 preservationLevel (archivierungsLevel)	
1.3.1 preservationLevelValue (archivierungsLevelWert)	full
1.5 objectCharacteristics (ObjektMerkmale)	
1.5.1 compositonLevel (kodierungsLevel)	0
1.5.2 fixity (datenIntegritaet)	
1.5.2.1 messageDigestAlgorithm (pruefsummeAlgorithmus)	MD5
1.5.2.2 messageDigest (pruefsummeWert)	36b03197ad066cd719906c55eb68ab8d
1.5.2.3 messageDigestOriginator (pruefsummeErzeuger)	localDCMS
1.5.3 size (datenmenge)	20800896
1.5.4 format (format)	
1.5.4.1 formatDesignation (formatBezeichnung)	
1.5.4.1.1 formatName (formatName)	image/tiff
1.5.4.1.2 formatVersion (formatVersion)	01.06.00
1.5.4.2 formatRegistry (formatVerzeichnis)	
1.5.4.2.1 formatRegistryName (formatVerzeichnisName)	PRONOM
1.5.4.2.2 formatRegistryKey (formatVerzeichnisIdentifikator)	fmt/10
1.5.4.2.3 formatRegistryRole (formatVerzeichnisRolle)	Spezifikation
1.5.5 creatingApplication (erzeugendesProgramm)	
1.5.5.1 creatingApplicationName (erzeugendesProgrammName)	ScandAll 21
1.5.5.2 creatingApplicationVersion (erzeugendesProgrammVersion)	4.1.4
1.5.5.3 dateCreatedByApplication (erzeugtVonProgrammAm)	1998-10-30
1.5.5 creatingApplication (erzeugendesProgramm)	
1.5.5.1 creatingApplicationName (erzeugendesProgrammName)	Adobe Photoshop
1.5.5.2 creatingApplicationVersion (erzeugendesProgrammVersion)	CS2
1.5.5.3 dateCreatedByApplication (erzeugtVonProgrammAm)	1998-10-30T08:29:02
1.5.6 inhibitors (zugriffssperren)	
1.5.6.1 inhibitorType (zugriffssperrenTyp)	keine
1.6 originalName (originalName)	001h.tif
1.7 storage (speicherung)	
1.7.1 contentLocation (speicherOrt)	
1.7.1.1 contentLocationType (speicherOrtTyp)	Dateipfad

1.7.1.2 contentLocationValue (speicherOrtWert)	amserver/
1.7.2 storageMedium (speicherMedium)	IBM DS4000 System 1740-5208
1.8 environment (technischeUmgebung)	
1.8.1 environmentCharacteristic (technischeUmgebungMerkmal)	Empfohlen
1.8.2 environmentPurpose (technischeUmgebungZweck)	Rendering
1.8.2 environmentPurpose (technischeUmgebungZweck)	Editieren
1.8.5 software (software)	
1.8.5.1 swName (swName)	Adobe Acrobat
1.8.5.2 swVersion (swVersion)	5.0
1.8.5.3 swType (swTyp)	Renderer
1.8.5.1 swName (swName)	Windows
1.8.5.2 swVersion (swVersion)	XP
1.8.5.3 swType (swTyp)	Betriebssystem
1.8.6 hardware (hardware)	
1.8.6.1 hwName (hwName)	Intel x86
1.8.6.2 hwType (hwTyp)	Prozessor
1.8.6.3 hwOtherInformation (hwWeitereInformation)	60 MHz Minimum
1.8.6 hardware (hardware)	
1.8.6.1 hwName (hwName)	64 MB RAM
1.8.6.2 hwType (hwTyp)	Arbeitsspeicher
1.8.6.3 hwOtherInformation (hwWeitereInformation)	32 MB minimum
1.10 relationship (objektBeziehungen)	
1.10.1 relationshipType (objektBeziehungenTyp)	Strukturell
1.10.2 relationshipSubType (objektBeziehungenUnterTyp)	Ist verwandt mit
1.10.3 relatedObjectIdentification (beziehungsObjektIdentifikator)	
1.10.3.1 relatedObjectIdentifierType (beziehungsObjektIdentifikatorTyp)	hdl
1.10.3.2 relatedObjectIdentifierValue (beziehungsObjektIdentifikatorWert)	loc.music/gottlieb.09602
1.10.3.3 relatedObjectSequence (beziehungsObjektSequenz)	0
1.10 relationship (objektBeziehungen)	
1.10.1 relationshipType (objektBeziehungenTyp)	Strukturell
1.10.2 relationshipSubType (objektBeziehungenUnterTyp)	Ist verwandt mit
1.10.3 relatedObjectIdentification (beziehungsObjektIdentifikator)	
1.10.3.1 relatedObjectIdentifierType (beziehungsObjektIdentifikatorTyp)	hdl
1.10.3.2 relatedObjectIdentifierValue (beziehungsObjektIdentifikatorWert)	loc.music/gottlieb.mets
1.10.3.3 relatedObjectSequence (beziehungsObjektSequenz)	0

1.10 relationship (objektBeziehungen)	
1.10.1 relationshipType (objektBeziehungenTyp)	strukturell
1.10.2 relationshipSubType (objektBeziehungenUnterTyp)	Ist enthalten in
1.10.3 relatedObjectIdentification (beziehungsObjektIdentifikator)	
1.10.3.1 relatedObjectIdentifierType (beziehungsObjektIdentifikatorTyp)	LokalesLangzeitarchiv
1.10.3.2 relatedObjectIdentifierValue (beziehungsObjektIdentifikatorWert)	R200802948
1.10.3.3 relatedObjectSequence (beziehungsObjektSequenz)	0
1.12 linkingIntellectualEntityIdentifier (verknuepftelIntellektuelleEntitaetIdentifikator)	
1.12.1 linkingIntellectualEntityIdentifierType (verknuepftelIntellektuelleEntitaetIdentifikatorTyp)	LCCN
1.12.2 linkingIntellectualEntityIdentifierValue (verknuepftelIntellektuelleEntitaetIdentifikatorWert)	2007-86121

Appendix C: Glossar der Begriffe

Dieses Glossar führt die Definitionen, die sich im vorangegangenen Text des Leitfadens finden, zusammen. Diese Definitionen können weniger formal sein als die des PREMIS Data Dictionary für Langzeitarchivierungsmetadaten.

Agent: Eine Person, Organisation oder ein Computerprogramm, die/ das im Zusammenhang mit einem *Ereignis* oder Angaben zu *Rechten* eine Rolle spielt.

ausführbar: Die Eigenschaft in einer kontrollierten Form verzeichnet zu sein, die eine Verarbeitung durch Computerprogramme ermöglicht.

Bitstream-Objekt: Ein Typ von PREMIS *Objekt*; Daten innerhalb einer Datei, die über gemeinsame, für die Zwecke der Langzeitarchivierung relevante Eigenschaften verfügen und nicht für sich alleine stehen können.

Containereinheiten: *Semantische Einheiten*, die selbst keinen Wert annehmen können, aber der Zusammenführung von in Beziehung zueinander stehenden Untereinheiten dienen.

Datei-Objekt: Ein Typ von PREMIS *Objekt*; eine Computerdatei wie PDF oder JPEG.

Digitale Provenienz: Dokumentation der Verarbeitungskette und der Veränderungshistorie einer digitalen Ressource.

Emulation: Eine *Langzeitarchivierungsstrategie*, die die Wiederherstellung einer alten Darstellungsumgebung auf einer neueren Hardware und/oder Software umfasst.

Entität: In PREMIS eine Art von Dingen, die beschrieben werden können. Die PREMIS Entitäten sind *Intellektuelle Entitäten*, *Objekte*, *Agenten*, *Ereignisse* und *Rechte*.

Entität Ereignis: Eine PREMIS *Entität*, für die Sammlung von Informationen über Handlungen, die die Objekte innerhalb eines Langzeitarchivs betreffen.

Entität Rechte: Ein PREMIS *Entität* für die Sammlung von Informationen über Rechte und Genehmigungen, die Objekte innerhalb eines Langzeitarchivs betreffen.

Erweiterungscontainer: Ein spezieller Typ von PREMIS *Containereinheit*, für den keine Untereinheiten definiert sind, der aber als Platzhalter für nicht-PREMIS Metadaten dient.

Intellektuelle Entität: Ein Set von Inhalten, das unter den Aspekten Management und Beschreibung als eine einzige intellektuelle Einheit betrachtet wird; ähnlich wie eine „bibliographische Entität“ in der Bibliothekswissenschaft.

Langzeitarchivierungsmetadaten: Metadaten, die Maßnahmen unterstützen, die die Sicherstellung der Langzeitverfügbarkeit von digitalen Ressourcen zum Ziel haben.

Langzeiterhaltungsstrategien: Techniken, die eingesetzt werden um sicherzustellen, dass digitale Objekte dauerhaft benutzbar bleiben; zwei geläufige Strategien sind *Migration* und *Emulation*.

Migration: Eine *Langzeitarchivierungsstrategie*, die die Erzeugung einer Version einer Datei in einem neueren Dateiformat umfasst.

Objekte: Digitale Elemente, die tatsächlich in einem digitalen Langzeitarchiv gespeichert und gemanagt werden. In PREMIS sind drei Typen von Objekten definiert: *Datei*, *Bitstream* und

Repräsentation.

Repräsentation: Ein Typ von PREMIS *Objekt*; ein Set von allen *Datei-Objekten*, die für die Darstellung bzw. die Anwendung einer intellektuellen Entität benötigt werden.

Semantische Einheiten: Ein Stück Information oder Wissen.

Signifikante Eigenschaften: Charakteristika eines Objekts, die durch Langzeitarchivierungsmaßnahmen erhalten werden sollen.

Technische Umgebung: Die Hard- und Software sowie andere Komponenten, die für die Darstellung oder Anwendung eines Objekts benötigt werden.

Zugriffssperren: Merkmale eines digitalen Objekts, die den Zugang, die Nutzung oder die *Migration* einschränken sollen.