

Library of Congress Network Development and MARC Standards Office

# PREMIS VERSTEHEN (2017)

## **Understanding PREMIS**

**Autorin: Priscilla Caplan**

**Überarbeitung: PREMIS-Herausgeberkomitee**

2009 erstmals veröffentlicht; 2017 überarbeitet  
Deutsche Übersetzung im Auftrag von nestor

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

Die Übersetzung basiert auf einer Arbeit von Tobias Beinert, Bayerische Staatsbibliothek, im Rahmen eines Projekts des Masterstudiengangs Bibliotheks- und Informationswissenschaft der Fachhochschule Köln und wurde ergänzt und aktualisiert von Proverb Ohg.

Im Auftrag von nestor – Kompetenznetzwerk Langzeitarchivierung und  
Langzeitverfügbarkeit digitaler Ressourcen für Deutschland



# Inhalt

|   |    |
|---|----|
| <b>Inhalt</b> .....   | 2  |
| <b>PREMIS VERSTEHEN</b> .....   | 3  |
| <b>1. DER ANWENDUNGSKONTEXT VON PREMIS</b> .....                                    | 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  |
| <b>2. KONVENTIONEN DES DATA DICTIONARY</b> .....                                    | 7  |
| 2.1. Semantische Einheiten .....  | 7  |
| 2.2. Container und untergeordnete Einheiten .....                                   | 7  |
| 2.3. Erweiterungscontainer .....  | 8  |
| <b>3. DAS PREMIS DATENMODELL</b> .....  | 9  |
| 3.1. Entität Objekt .....   | 9  |
| 3.2. Ereignisse .....   | 11 |
| 3.4. Agenten .....  | 12 |
| 3.4. Rechte .....   | 12 |
| <b>4. DAS DATA DICTIONARY</b> .....   | 13 |
| 4.1. Beispieleintrag im Data Dictionary für eine einfache semantische Einheit ..... | 13 |
| 4.2. Beispieleintrag im Data Dictionary für eine Container-Einheit .....            | 15 |
| <b>5. DIE ANWENDUNG VON PREMIS</b> .....  | 16 |
| 5.1. PREMIS in XML .....  | 16 |
| 5.2. PREMIS Konformität .....   | 16 |
| <b>6. WEITERE INFORMATIONEN</b> .....   | 18 |
| <b>Appendix A: Beispiele</b> .....  | 19 |
| A.1. Objektbeispiel .....   | 19 |
| A.2. Ereignisbeispiel .....   | 21 |
| A.3. Agentenbeispiel .....  | 23 |
| <b>Appendix B: Glossar der Begriffe</b> .....                                       | 24 |

# 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 mehr als 250-seitige 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.

## 1. DER ANWENDUNGSKONTEXT VON PREMIS

### 1.1. Was sind Langzeitarchivierungsmetadaten?

Wenn Sie in einer Bibliothek, einem Archiv oder Museum 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, sodass diese ordnungsgemäß gerendert 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. Dies könnte bedeuten, dass alte Formate auf neuere Äquivalente migriert werden müssen, oder dass die alte Rendering-Umgebung auf neuerer Hardware und Software emuliert werden muss. Sowohl für Migration als auch für Emulation werden Metadaten zum ursprünglichen Dateiformat und zur Unterstützung durch Hardware- und Softwareumgebungen benötigt.
- Langzeitarchivierungsstrategien machen es teilweise nötig, ursprüngliche Ressourcen zu verändern (Migration) oder ihr Rendering zu verändern (Emulation). 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 "**PRE**servation **Meta**data: **Im**plementation **Str**ategies", 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. Die Library of Congress veröffentlichte daraufhin eine Reihe von PREMIS-Schemata zur Darstellung von Metadatenelementen im Data Dictionary in XML. Eine überarbeitete zweite Ausgabe des Data Dictionary und ein neues unterstützendes Schema wurden im März 2008 veröffentlicht, gefolgt von kleinen Änderungen die Versionen 2.1 und 2.2. Im Juni 2015 wurde eine umfassendere Überarbeitung, Version 3.0, veröffentlicht.

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 für Überarbeitungen des Data Dictionary und der Schemata verantwortlich ist. Die Maintenance Activity versucht darüber hinaus den Bekanntheitsgrad von PREMIS zu steigern, unterstützt Tutorials und Implementierungs-Meetings 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 generell 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: <http://www.loc.gov/premis/v3/premis-3-0-final.pdf>

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

PREMIS Anwendergruppe Diskussionsliste: [pig@loc.gov](mailto:pig@loc.gov)

Um sich anzumelden, senden Sie eine E-Mail an [mlistserv@loc.gov](mailto:mlistserv@loc.gov) mit folgender Nachricht: subscribe pig [Ihr Name]

## 1.3. Was steht im PREMIS Data Dictionary?

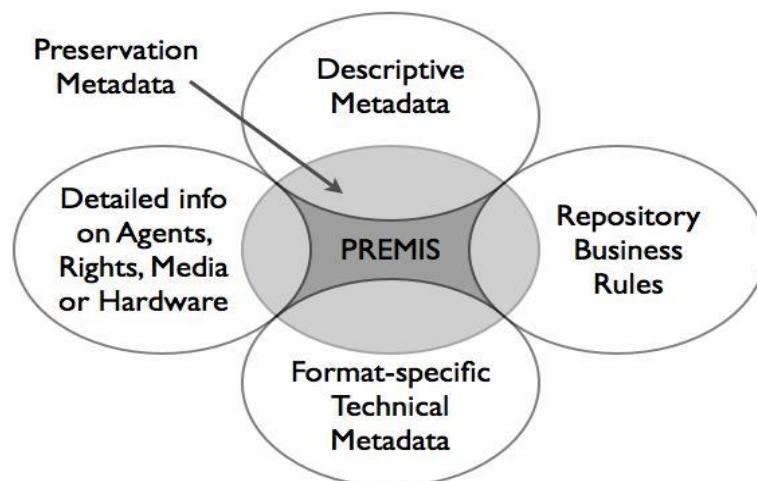
Im PREMIS Data Dictionary wird ein Kernset an sogenannten Metadatenelementen definiert (eigentlich „semantische Einheiten“ (semantic units), doch dazu später mehr), die Langzeitarchiven bekannt sein sollten, damit sie ihre Archivierungsfunktionen erfüllen können. Wenngleich Art und Umfang der Langzeitarchivierung 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, können zu diesem Zweck eine Vielzahl an unabhängigen Standards eingesetzt werden, einschließlich MARC, MODS und Dublin Core.
- Detaillierte Informationen über Speichermedien oder Hardware. Auch hier gilt, dass obwohl diese offensichtlich relevant für die Langzeitarchivierung sind, es anderen Communities überlassen ist, Metadaten dafür festzulegen.
- Detaillierte Informationen über Agenten (Personen, Organisationen oder Software), die über zur Identifizierung notwendige Informationen hinausgehen.
- Umfangreiche Informationen über Rechte und Genehmigungen; der Fokus liegt auf jenen, die sich auf Langzeitarchivierungsfunktionen 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 Langzeitarchivierungsmetadaten als Kreis in der Mitte des Diagramms. Er enthält einige deskriptive Metadaten, einige Geschäftsregeln, einige detaillierte technische Metadaten und einige detaillierte Informationen über Agenten, Rechte, Medien und Hardware. PREMIS bildet den kleinen Kern in der Mitte der Langzeitarchivierungsmetadaten, der all diese anderen 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 anderweitig 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 Archivinformationspaketen 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, ein Kunde könnte von einem externen Serviceanbieter zu einem anderen wechseln wollen oder eine Institution könnte den Langzeitarchivierungsservice einer anderen Institution in einem vertrauenswürdigen digitalen Langzeitarchiv nutzen. 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. Ziehen Sie in Betracht, die Schulungsmaterialien auf der PREMIS-Webseite zu nutzen oder an einem PREMIS-Tutorial teilzunehmen, wenn diese angeboten werden.

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 derzeit allen 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 Zugänglichkeit, die Benutzung oder die Migration verhindern sollen. Zu den Zugriffssperren gehören der Schutz durch Passwörter und die Verschlüsselung. Langzeitarchivsoftware kann eine Zugriffssperre womöglich nicht identifizieren, da diese eine Analyse des Objekts verhindert. Bekannte Zugriffssperren sollten dokumentiert werden. 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. Ein kontrolliertes Vokabular für Ereignistypen steht unter [id.loc.gov/vocabulary/preservation/eventType](http://id.loc.gov/vocabulary/preservation/eventType) zur Verfügung.

Viele der PREMIS-Ereignistypen sind in erster Linie dafür gedacht, Handlungen zu beschreiben, die nach der Abgabe von Objekten für die Übernahme in ein Langzeitarchiv durchgeführt werden, obwohl einige für Ereignisse vor der Übernahme, wie Auswahl und Erwerbung, verwendet werden.

**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. Dennoch sollten sich alle Institutionen, die für eine Nutzergruppe digitale Materialien erstellen oder erwerben, gründlich darüber Gedanken machen, welche Eigenschaften dieser Materialien für diese Nutzergruppe von Bedeutung sind und versuchen, diese Informationen als signifikante Eigenschaften 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. Alle bekannten rechtlichen Informationen, inklusive des urheberrechtlichen Status, der Lizenzbedingungen sowie spezieller Befugnisse, sollten aufgezeichnet werden.

## 2. KONVENTIONEN DES DATA DICTIONARY

### 2.1. Semantische Einheiten

Das PREMIS Data Dictionary definiert keine Metadatenelemente, sondern semantische Einheiten. Dies ist ein geringfügiger, aber realer 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 semantischen Einheiten denken. Für alle anderen kommen Metadatenelemente dem Ganzen auch sehr nahe.

Die Namen der semantischen Einheiten von PREMIS sind „camelCase“-Strings. Das heißt, Begriffe sind nicht durch einen Abstand voneinander getrennt, sondern beginnen mit Großbuchstaben: `objectIdentifier` (objektIdentifikator), `relatedEventIdentification` (zugehoerigesEreignis 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 das verpflichtende (M; mandatory) und wiederholbare (R; repeatable) Element `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 sind (NR; non-repeatable), müsste die gesamte Containerstruktur wiederholt werden, um zwei verschiedene Identifikatoren zu verzeichnen.

### 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, da dies als außerhalb des Geltungsbereichs erachtet würde, 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 Schemata, wie dem Data Dictionary Z39.97 für Bitmap Images definiert sind (deren XML-Schema wird als „NISO Metadata for Images in XML Schema--MIX“ bezeichnet).

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. Spätere Bemühungen brachten eine PREMIS OWL Ontology zur Nutzung für Linked-Data-Softwareanwendungen hervor. Die erste Version der Ontologie war kompatibel mit der PREMIS-Version 2.2 und zu diesem Zeitpunkt ist eine Überarbeitung für Version 3.0 fast fertig. Informationen über die PREMIS OWL Ontology sind unter [www.loc.gov/standards/premis/ontology](http://www.loc.gov/standards/premis/ontology) verfügbar.



### 3. DAS PREMIS DATENMODELL

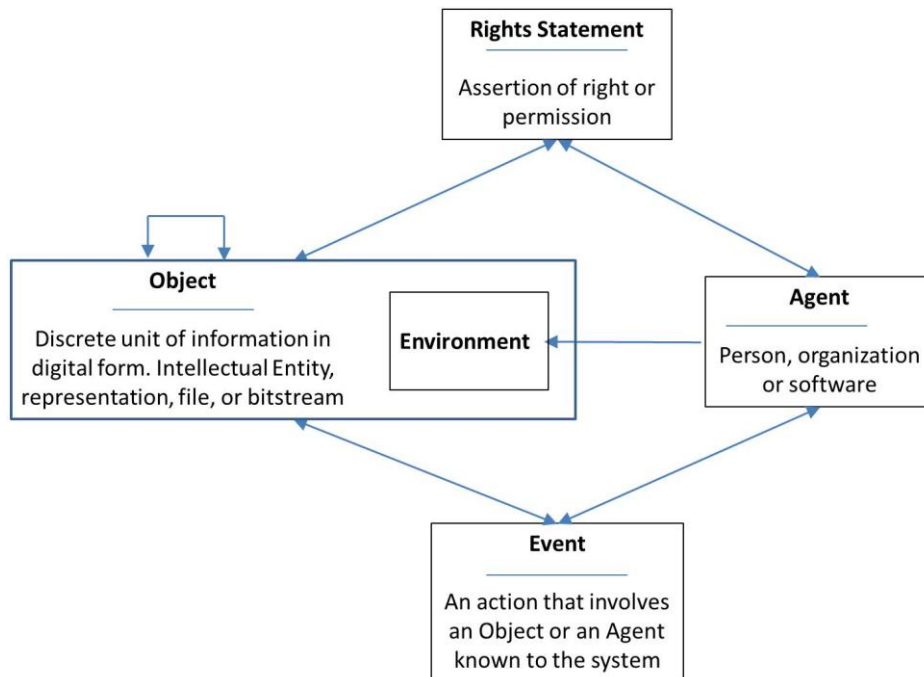


Abbildung 2: Das PREMIS Datenmodell, Version 3

Eines der Hauptprinzipien hinter PREMIS ist, dass man sich sehr klar darüber sein muss, was man beschreibt.

In den Versionen 1 und 2 definierte PREMIS fünf Arten von Dingen (sogenannte Entitäten), von denen man reden kann: Intellektuelle Entität, Objekt, Agent, Ereignis und Rechte. (Ein Diagramm des Datenmodells für die Versionen 1 und 2 liegt auf der PREMIS-Website vor.) In Version 3, die im Juni 2015 veröffentlicht wurde, wurde das PREMIS-Datenmodell überarbeitet, um die Intellektuelle Entität als neue Objektkategorie hinzuzufügen.

#### 3.1. 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.

In PREMIS sind vier unterschiedliche Typen von Objekten definiert und es ist erforderlich, dass die Anwender eine Unterscheidung zwischen diesen treffen. Es gibt Bitstreams (bitstreams), Dateien (Files), Repräsentationen (representations) und Intellektuelle Entitäten (Intellectual Entities).

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

**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.

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 der 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.

Das Objekt **Intellektuelle Entität** ist definiert 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 deskriptiven Metadaten, die Intellektuelle Entitäten betreffen, da bereits zahlreiche deskriptive Metadatenstandards zur Verfügung stehen. In Version 1 und 2 konnten Intellektuelle Entitäten nur mit einem Identifier referenziert werden, weil sie rein konzeptionell betrachtet wurden. In Version 3 kann eine Intellektuelle Entität aber mit deskriptiven Metadaten außerhalb von PREMIS oder mit Langzeitarchivierungsmetadaten als ein Objekt innerhalb von PREMIS beschrieben werden. In den meisten Fällen sind die semantischen Einheiten, die für Intellektuelle Entitäten verwendet werden, die gleichen wie für die Repräsentationen.

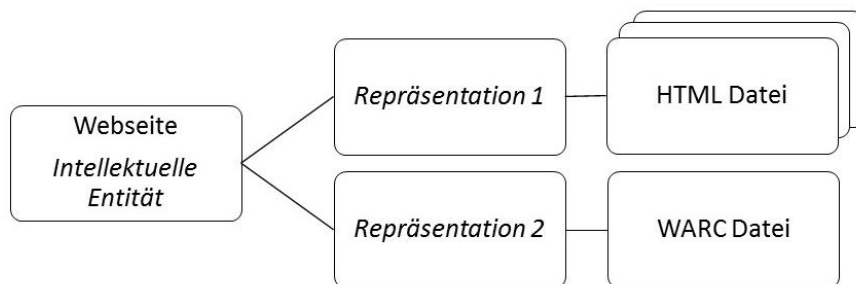
PREMIS gibt vor, dass ein Objekt in einem Archivierungssystem mit der konzeptuellen Intellektuellen Entität, die es repräsentiert, verknüpft sein soll, indem es einen Identifikator für die Intellektuelle Entität in die Metadaten des Objekts aufnimmt. Wenn man zum Beispiel eine Kopie des Buches *Buddhism: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

Einige der im PREMIS Data Dictionary definierten semantischen Einheiten sind auf alle vier Typen von Objekten anwendbar, während andere nur für einen oder zwei Typen von Objekten verwendet werden können.

Einige semantische Einheiten sind so definiert, dass sie die Umgebung eines Objekts verzeichnen, das heißt, welche Hardware und Software benötigt werden, um es zu rendern und welche Abhängigkeiten zu anderen Objekten es gibt. In den Versionen 1 und 2 sind diese Informationen Teil der Objektbeschreibung. In Version 3 kann zur Umgebung, als spezieller Art von Objekt der Intellektuellen Entität, von den Dateien, Repräsentationen und Bitstreams aus, die diese nutzen, verlinkt werden. Zum Beispiel kann eine PDF-Datei sowohl von einigen Versionen von Adobe Acrobat oder Adobe Reader angezeigt werden als auch von anderen Open Source- und kommerziellen Programmen. Jedes dieser wird wiederum auf mehreren Betriebssystemen

unterstützt und erfordert bestimmte minimale Hardware-Spezifikationen (Prozessor, Geschwindigkeit, Speicher und Festplatte). Weil Adobe Reader keine eigenständige Anwendung ist, sondern ein Browser-Plug-In, ist es auch abhängig von bestimmten Versionen bestimmter Browser; zum Beispiel benötigt Reader 9 für Mac OS den Safari Browser 2.0.4 oder später. Informationen über die Umgebung sind entscheidend für bestimmte Langzeitarchivierungsstrategien, aber die Bestimmung ist schwierig und zeitaufwändig und die Aufzeichnung erfolgt idealerweise in zentralen Registries (Verzeichnissen) wie dem PRONOM-Registry, das durch The National Archives des Vereinigten Königreichs gewartet wird. PREMIS erlaubt es Langzeitarchiven, zu Informationen, die in externen Registries vorliegen, zu verlinken, wenn dies einer lokalen Verzeichnung vorzuziehen ist.

Um die Beziehungen zwischen Objekten zu veranschaulichen, empfehlen wir einen erneuten Besuch der oben genannten Website. Die Website ist eine *Intellektuelle Entität*, die in einer Reihe verschiedener *Repräsentationen* ausgedrückt werden kann. Eine Repräsentation besteht, wie oben aufgeführt, aus einer Reihe unterschiedlicher Dateiobjekte (HTML-Seite, Bilder, Stylesheet etc.). Die Website kann jedoch auch als einzelne Web-Archiv-Datei (WARC) gespeichert werden (siehe [www.digitalpreservation.gov/formats/fdd/fdd000236.shtml](http://www.digitalpreservation.gov/formats/fdd/fdd000236.shtml) für weitere Informationen). Dies ist eine unterschiedliche Repräsentation der gleichen Intellektuellen Entität, wobei diese, wie dargestellt, nur eine einzige Datei enthält:



**Abbildung 3: PREMIS Beispiel von Objektbeziehungen**

### 3.2. 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, und stellt ein kontrolliertes Vokabular wichtiger Ereignistypen zur Verfügung, um Langzeitarchive dabei zu unterstützen, diese Ereignisse einheitlich zu verzeichnen.

Dieses Vokabular ist unter <http://id.loc.gov/vocabulary/preservation/eventType> verfügbar.

### 3.4. Agenten

*Agenten* sind Handelnde, die bestimmte Rollen bei Ereignissen und in Bezug auf rechtlichen Angaben (siehe 3.4. Rechte) und Umgebungsobjekte haben. Agenten können Personen, Organisationen, Softwareanwendungen oder Hardware sein. PREMIS definiert nur eine Minimalzahl an notwendigen semantischen Einheiten für die Bestimmung von Agenten, da es bereits zahlreiche externe Standards gibt, die für genauere Informationen verwendet werden können. 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 umfasst:

- einen eindeutigen Identifikator für den Agenten (Typ und Wert),
- den Namen des Agenten,
- Bestimmung der Art des Agenten (Person, Organisation, Software, Hardware)
- Agentenversion (für Software oder Hardware),
- einen allgemeinen Hinweis zum Agenten,
- Ereignisse, die mit dem Agenten in Verbindung stehen,
- rechtliche Angaben, die mit dem Agenten in Verbindung stehen,
- Umgebungsobjekte, die mit dem Agenten in Verbindung stehen

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.

### 3.4. Rechte

Die meisten Langzeitarchivierungsstrategien umfassen das Anfertigen von identischen Kopien und abgeleiteten Versionen digitaler Objekte – Handlungen, die laut Urheberrecht den Rechteinhabern vorbehalten sein könnten. Die Entität *Rechte* fasst Informationen über Rechte und Erlaubnisse bezüglich Objekten in einem Langzeitarchiv zusammen, sodass das Langzeitarchiv alles Erforderliche zu deren Erhaltung vornehmen kann. 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, ein Statut oder eine andere gesetzliche Bestimmung ist (z. B. institutionelle Politik),
- 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 Beschränkung, oder den Zeitraum, für den die Angabe Anwendung findet
- 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/](http://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. Beispielintrag im Data Dictionary für eine einfache semantische Einheit

Tabelle 1 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. Falls ein kontrolliertes Vokabular vorliegt, wird darauf in der Datenbeschränkung verwiesen, und einige Begriffe im Vokabular sind als Beispiele im Data Dictionary spezifiziert: <http://id.loc.gov/preservationdescriptions/>

Es können andere kontrollierte Vokabulare genutzt werden; in diesem Fall müssen die Namen der Vokabulare erfasst werden. Im Data Dictionary sind keine semantischen Einheiten für die Namen von Vokabularen definiert, aber das PREMIS-XML-Schema stellt einen Ort für sie zur Verfügung, falls sie von einer URI identifiziert werden, sodass generell der Name des kontrollierten Vokabulars widergespiegelt wird.

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 Langzeitarchiv wissen, dass es sich bei seinen Dateien um TIFF 6.0 Images handelt und könnte diese Information weitergeben, wenn es müsste. (siehe Kapitel 5.2 PREMIS-Konformität). Einige semantische Einheiten sind innerhalb eines Containers verpflichtend; falls dieser Container nicht verpflichtend ist, wird die semantische Einheit nicht verwendet, wenn der Container abwesend ist.

**Tabelle 1: Auszug aus dem Data Dictionary für die semantische Einheit size**

|   |   |  |  |
|---|---|--|--|
| <b>Semantic Unit<br/>(Semantische Einheit)</b>  | 1.5.3. size<br>(1.5.3. datenMenge)  |  |  |
| <b>Semantic Components<br/>(Semantische Komponenten)</b>  | None<br>(Keine)   |  |  |
| <b>Definition<br/>(Definition)</b>  | Die Größe der im Langzeitarchiv gespeicherten Datei oder des Bitstreams in Bytes  |  |  |
| <b>Rationale</b>  | Die Größe ist nützlich, um zu gewährleisten, dass aus dem Speicher 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 Abrechnung von Speicherkosten verwendet werden. |  |  |
| <b>Data constraint<br/>(Datenbeschränkung)</b>  | Integer<br>(Ganze Zahlen)   |  |  |
| <b>Object Category<br/>(Objektkategorie)</b>  | Repräsentation  | Datei                                  | Bitstream                              |
| <b>Applicability<br/>(Anwendbarkeit)</b>  | Not applicable<br>(Nicht anwendbar)   | Applicable<br>(Anwendbar)              | Applicable<br>(Anwendbar)              |
| <b>Examples<br/>(Beispiele)</b>   |   | 2038937                                |  |
| <b>Repeatability<br/>(Wiederholbarkeit)</b>   |   | Not repeatable<br>(Nicht wiederholbar) | Not repeatable<br>(Nicht wiederholbar) |
| <b>Obligation<br/>(Verpflichtung)</b>   |   | Optional<br>(optional)                 | Optional<br>(optional)                 |
| <b>Creation /<br/>Maintenance Notes<br/>(Bemerkungen zur<br/>Entstehung / Erhaltung<br/>und Pflege)</b> | Automatisch vom Langzeitarchivierungssystem aufgezeichnet.  |  |  |
| <b>Usage Notes<br/>(Bemerkungen zum<br/>Gebrauch)</b>   | 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.  |  |  |

## 4.2. Beispieleintrag im Data Dictionary für eine Container-Einheit

Tabelle 2 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.

**Tabelle 2: Auszug aus dem Data Dictionary für die semantische Einheit objectCharacteristics**

|   |   |                              |                              |
|---|---|------------------------------|------------------------------|
| <b>Semantic Unit<br/>(Semantische Einheit)</b>              | 1.5. objectCharacteristics (objektMerkmale)   |                              |                              |
| <b>Semantic Components<br/>(Sematische<br/>Komponenten)</b> | 1.5.1 compositionLevel (kodierungsLevel)<br>1.5.2 fixity (datenIntegritaet)<br>1.5.3 size (Datenmenge)<br>1.5.4 format (format)<br>1.5.5 creatingApplication (erzeugendesProgramm)<br>1.5.6 inhibitors (Zugriffssperren)<br>1.5.7 objectCharacteristicsExtension (objektMerkmaleErweiterung)  |                              |                              |
| <b>Definition<br/>(Definition)</b>                          | Technische Eigenschaften einer Datei oder eines Bitstreams, die auf alle oder die meisten Formate zutreffen.  |                              |                              |
| <b>Rationale<br/>(Begründung)</b>                           | 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.  |                              |                              |
| <b>Data constraint<br/>(Datenbeschränkung)</b>              | Container<br>(Container)  |                              |                              |
| <b>Object Category<br/>(Objektkategorie)</b>                | Intellektuelle Entität /<br>Repräsentation  | Datei                        | Bitstream                    |
| <b>Applicability<br/>(Anwendbarkeit)</b>                    | Not applicable<br>(Nicht anwendbar)   | Applicable<br>(Anwendbar)    | Applicable<br>(Anwendbar)    |
| <b>Repeatability<br/>(Wiederholbarkeit)</b>                 |   | Repeatable<br>(Wiederholbar) | Repeatable<br>(Wiederholbar) |
| <b>Obligation<br/>(Verpflichtung)</b>                       |   | Mandatory<br>(Verpflichtend) | Mandatory<br>(Verpflichtend) |
| <b>Usage Notes<br/>(Anmerkungen zum<br/>Gebrauch)</b>       | Die semantischen Einheiten, die in objectCharacteristics enthalten sind, sollten als ein Set von Informationen verwendet werden, das ein einzelnes Objekt auf einem einzigen compositionLevel betrifft. ObjectCharacteristics kann wiederholt werden, wenn ein Objekt durch die Anwendung von zwei oder mehr Kodierungen, wie zum Beispiel Komprimierung und Verschlüsselung, erstellt wurde. In diesem Fall hätte jede Wiederholung von objectCharacteristics ein zunehmend höheres compositionLevel. Wird die Verschlüsselung angewandt, muss der Block objectCharacteristics eine semantische Sperreinheit beinhalten. Ein in eine Datei eingebetteter Bitstream kann unterschiedliche objectCharacteristics als die Datei haben. Wo diese Eigenschaften für die Langzeitarchivierung relevant sind, sollten sie verzeichnet werden. |                              |                              |

## 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 4 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) – ein weiterer von der Library of Congress betreuter 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 ein Feld 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 jede Implementierung diesbezüglich seine eigenen Entscheidungen trifft, es zu einer großen Bandbreite in der Repräsentation der Daten kommen kann, die die Interoperabilität beeinträchtigt. Deshalb haben Anwender Best Practice-Beispiele für einen gemeinsamen Einsatz von PREMIS und METS entwickelt. Richtlinien für die Nutzung von PREMIS mit METS zum Austausch sind auf der Website der PREMIS Maintenance Activity verfügbar:

<http://www.loc.gov/standards/premis/guidelines2017-premismets.pdf>

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

Abbildung 4: Ein PREMIS-Auszug in XML

### 5.2. PREMIS Konformität

Die PREMIS-Spezifikation enthält einen Abschnitt darüber, was es für ein Langzeitarchiv bedeutet, PREMIS-konform zu sein. Im Grunde spezifiziert das Konformitäts-Statement, welches das PREMIS-Herausgeberkomitee erstmals im Oktober veröffentlichte und 2015 überarbeitete, Anwendungsprinzipien, Konformitätslevel und einen Implementierungsleitfaden.

Er ist verfügbar unter: <http://www.loc.gov/standards/premis/premis-conformance-20150429.pdf>.



## Anwendungsprinzipien

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. Wenn ein Metdatenelement die Definition einer semantischen Einheit von PREMIS teilt, aber nicht deren Namen, muss das Langzeitverzeichnis ein Mapping zwischen dem Metdatenelement und dessen zugehöriger semantischer Einheit von PREMIS herstellen.

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) Eine Implementierung muss die verpflichtenden semantischen Einheiten für alle Datenmodellentitäten (Objekt, Ereignis, Agent oder Rechte) enthalten, die vom Langzeitarchiv unterstützt werden. Es ist zu beachten, dass eine verpflichtende semantische Komponente nur enthalten ist, falls der Muttercontainer implementiert ist.

4) 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.

## Konformitätslevel

Die Level spezifizieren drei Arten der Aufrechterhaltung der Konformität mit PREMIS in einem Archivierungssystem:

- **Level 1.** Fähigkeit, Langzeitarchivierungsmetadaten zu PREMIS zu mappen
- **Level 2.** Fähigkeit, Langzeitarchivierungsmetadaten als PREMIS zu exportieren
- **Level 3.** Nutzung von PREMIS als internes Schema auf eine Weise, die kein weiteres Mapping oder eine Umwandlung erfordert.

Diese Level sind wiederum in Kategorien aufgeteilt: nur Implementierung der Objektentität oder Implementierung der Objektentität sowie einer weiteren oder mehrerer weiterer Entitäten. Mit anderen Worten muss ein Langzeitarchiv mindestens Informationen über das Objekt implementieren, es ist jedoch nicht erforderlich, dass es alle Entitätstypen, die im Datenmodell von PREMIS definiert sind, unterstützt. 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 anderen Wege. Solange also ein Langzeitarchiv einen guten PREMIS-Wert liefern kann, wenn dieser benötigt wird, ist es konform.

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

Die Website der PREMIS Maintenance Activity ([www.loc.gov/standards/premis/](http://www.loc.gov/standards/premis/)) bietet für alle etwas, einschließlich Links zu der PREMIS-Anwendergruppe (PREMIS Implementers Group – PIG), PREMIS-Implementierungsmessen sowie Tutorials, Schemata, Tools und Neuigkeiten. Es gibt außerdem die Rubrik PREMIS-Ressourcen („PREMIS Resources“), die Links zu Literatur und verwandten Themen bietet ([www.loc.gov/standards/premis/bibliography.html](http://www.loc.gov/standards/premis/bibliography.html)). Einige hilfreiche und allgemein verständliche Ressourcen sind hier aufgelistet:

Zu Langzeitarchivierungsmetadaten allgemein:

“Preservation Metadata” 2. Auflage (PDF:986KB/36pp.)  
Brian Lavoie (OCLC) and Richard Gartner (Oxford)  
Published by the Digital Preservation Coalition  
as DPC Technology Watch Report No. 13-03: May 2013.  
[dx.doi.org/10.7207/twr13-03](https://dx.doi.org/10.7207/twr13-03)

Zur Umsetzung von PREMIS

Digital Preservation Metadata for Practitioners: Implementing PREMIS.  
Angela Dappert, Rebecca Squire Guenther, Sébastien Peyrard, Editors. Springer, 2016.  
[dx.doi.org/10.1007/978-3-319-43763-7](https://dx.doi.org/10.1007/978-3-319-43763-7)

Zur Änderungen in PREMIS 3.0 (Webinar):

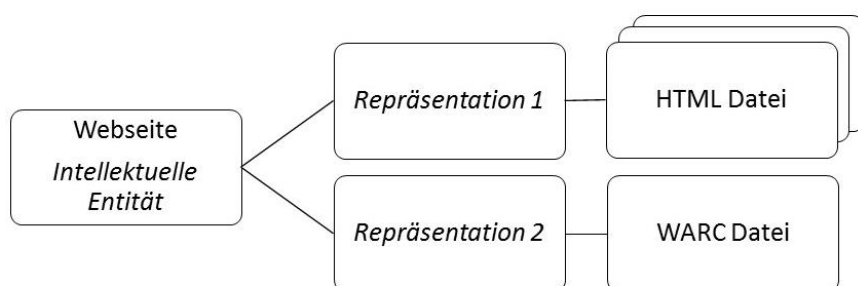
“Digital Preservation Metadata and Improvements to PREMIS in Version 3.0:  
A DCMI/ASIST Joint Webinar Presented by Angela Dappert”  
(Wednesday, May 27, 2015)  
<http://www.loc.gov/standards/premis/v3/tutorial.html>

## Appendix A: Beispiele

### A.1. Objektbeispiel

Dieses Beispiel beschreibt, wie eine Website in einem hypothetischen Langzeitarchivierungssystem modelliert werden könnte. Das Beispiel führt durch ein High-Level-Diagramm, liefert einige Details darüber, wie semantische Einheiten zum Beispiel bestückt werden können, und am Ende einige XML-Ausschnitte, um zu veranschaulichen, wie die semantischen Einheiten im Langzeitarchiv implementiert werden könnten. Die Beispiele zeigen nicht alle semantischen Einheiten, die für jedes Level des Objekts relevant sind.

Bei diesem Beispiel heißt die Website „PREMIS Website“. Das ist die Intellektuelle Entität, die langzeitarchiviert werden muss. Es gibt zwei Repräsentationen dieser Website. Für dieses Beispiel ist eine ein Langzeitarchivierungsmaster, die andere ein Zugangsmaster.



### Intellektuelle Entität

Das Level Intellektuelle Entität bei diesem Beispiel stellt eine wichtige Eigenschaft der Website dar. Es regelt das Verhalten beider Repräsentationen. Es ist natürlich viel wahrscheinlicher, dass wesentliche Eigenschaften viel detaillierter wären, aber das Beispiel dient zur Veranschaulichung, wie sie auf dem IE-Level verwendet werden können.

| PREMIS semantische Einheit       | Intellektuelle Entität 1                    |
|----------------------------------|---|
| 1.1 objectIdentifier             |   |
| 1.1.1 objectIdentifierType       | IEPID                                       |
| 1.1.1 objectIdentifierValue      | 17415491                                    |
| 1.4 significantProperties        |   |
| 1.4.1 significantPropertiesType  | Verhalten                                   |
| 1.4.2 significantPropertiesValue | Externe Links öffnen in einem neuen Fenster |

### Repräsentation

Das Langzeitarchivierungslevel beschreibt die institutionelle Richtlinie für dieses Objekt. Das Langzeitarchivierungslevel kann auf dem Level IE, Repräsentation oder Datei aufsetzen. In diesem Fall wurde es auf dem Level Repräsentation aufgesetzt, da die Institution wünscht, dass jede Repräsentation unterschiedlich gepflegt wird. Die Institution hat bei diesem Beispiel entschieden, dass die Repräsentation, die die WARC-Datei umfasst (Repräsentation 2) diejenige ist, die als Langzeitarchivierungsmaster gilt; also diejenige, die dauerhaft erhalten bleibt und von der aus Zugangsmaster erstellt werden. Repräsentation 1 ist der Zugangsmaster mit dem Zweck, dass er die Repräsentation ist, die der Öffentlichkeit zur Ansicht dient. Sein Wert ist nicht langfristig, sondern er bleibt solange erhalten, wie der aktuelle Zugangsmechanismus in Kraft ist.

**Repräsentation 1** wurde ein Wert „bit preservation“ zugeordnet. Die institutionelle Policy definiert, dass dafür nur Sicherheitskopien erstellt werden und keine Risikoanalyse oder daraus resultierende Emulation oder Migration erforderlich ist.

**Repräsentation 2** wurde dagegen als logische Langzeitarchivierung („logical preservation“) mit voller Leistungsfähigkeit klassifiziert. Dies wird von der Institution als Repräsentation definiert, bei der die relevante Risikoanalyse erfolgt, und wenn erforderlich, Emulation oder Migration genutzt wird, um die Langzeitarchivierung in der Zukunft zu gewährleisten.

| PREMIS semantische Einheit          | Repräsentation 1     | Repräsentation 2     |
|-------------------------------------|----------------------|----------------------|
| 1.1 objectIdentifier                |                      |                      |
| 1.1.1 objectIdentifierType          | REPPID               | REPPID               |
| 1.1.1 objectIdentifierValue         | 17415492             | 17415495             |
| 1.3 preservationLevel               |                      |                      |
| 1.3.1 preservationLevelType         | bit preservation     | logical preservation |
| 1.3.2 preservationLevelValue        | low                  | full                 |
| 1.3.3 preservationLevelRole         | capability           | intention            |
| 1.3.4 preservationLevelRationale    | institutional policy | institutional policy |
| 1.3.5 preservationLevelDateAssigned | 2015-02-23           | 2015-02-23           |
| 1.13 relationship                   |                      |                      |
| 1.13.1 relationshipType             | derivation           | derivation           |
| 1.13.2 relationshipSubType          | has source           | is source of         |
| 1.13.3 relatedObjectIdentifier      |                      |                      |
| 1.14.1 relatedObjectIdentifierType  | REPPID               | REPPID               |
| 1.14.2 relatedObjectIdentifierValue | 17415495             | 17415492             |

Ein XML-Ausschnitt für das Langzeitarchivierungslevel könnte so aussehen.

```

<premis:preservationLevel>
  <premis:preservationLevelType>logical preservation</premis:preservationLevelType>
  <premis:preservationLevelValue>full</premis:preservationLevelValue> <premis:preservationLevelRole
  authority="preservationLevelRole"
  authorityURI="http://id.loc.gov/vocabulary/preservation/preservationLevelRole"
  valueURI="http://id.loc.gov/vocabulary/preservation/preservationLevelRole/int">intention
</premis:preservationLevelRole>
  <premis:preservationLevelRationale>institutional policy</premis:preservationLevelRationale>
  <premis:preservationLevelDateAssigned>2015-02-23</premis:preservationLevelDateAssigned>
</premis:preservationLevel>

```

## Dateien

Repräsentation 1 enthält eine Reihe unterschiedlicher Dateien, von denen jede als Dateiobjekt mit ihrem eigenen Set von Metadatenelementen beschrieben wird.

| PREMIS Semantische Einheit      | Datei 1  | Datei 2  |
|---------------------------------|--|--|
| 1.1 objectIdentifier            |  |  |
| 1.1.1 objectIdentifierType      | FILEPID  | FILEPID  |
| 1.1.1 objectIdentifierValue     | 17415493   | 17415494   |
| 1.5 objectCharacteristics       |  |  |
| 1.5.1 compositionLevel          | 0  | 0  |
| 1.5.2 fixity                    |  |  |
| 1.5.2.1 messageDigestAlgorithm  | SHA256   | SHA256   |
| 1.5.2.2 messageDigest           | d2bed92b73c7090bb30a0b30016882e7069c437488e1513e9deaacbe29d38d92 | 074862dbfa0806ef5a26c3ff748d394e79728e9c957ff8c198ae13214c9cfec0 |
| 1.5.2.3 messageDigestOriginator | NRI  | NRI  |
| 1.5.3 size                      | 4859385  | 123451   |

|                                       |                                      |                              |
|---------------------------------------|--------------------------------------|------------------------------|
| 1.5.4 format                          |                                      |                              |
| 1.5.4.1 formatDesignation             |                                      |                              |
| 1.5.4.1.1 formatName                  | Extensible Hypertext Markup Language | JPEG File Interchange Format |
| 1.5.4.1.2 formatVersion               | 1                                    | 1.02                         |
| 1.5.4.2 formatRegistry                |                                      |                              |
| 1.5.4.2.1 formatRegistryName          | PRONOM                               | PRONOM                       |
| 1.5.4.2.2 formatRegistryKey           | fmt/102                              | fmt/44                       |
| 1.5.4.2.3 formatRegistryRole          | specification                        | specification                |
| 1.13 relationship                     |                                      |                              |
| 1.13.1 relationshipType               | structural                           | structural                   |
| 1.13.2 relationshipSubType            | Is included in                       | Is included in               |
| 1.13.3 relatedObjectIdentifier        |                                      |                              |
| 1.14.3.1 relatedObjectIdentifierType  | REPPID                               | REPID                        |
| 1.14.3.2 relatedObjectIdentifierValue | 17415492                             | 17415492                     |

Ein XML-Ausschnitt für eine Objektbeschreibung könnte wie folgt aussehen:

```

<premis:objectCharacteristics>
  <premis:compositionLevel>0</premis:compositionLevel>
  <premis:fixity>
    <premis:messageDigestAlgorithm>SHA256</premis:messageDigestAlgorithm>
    <premis:messageDigest>d2bed92b73c7090bb30a0b30016882e7069c437488e1513e9deaacbe29d38d92
      </premis:messageDigest>
    <premis:messageDigestOriginator> NRI</premis:messageDigestOriginator> </premis:fixity>
  <premis:size>4859385</premis:size>
  <premis:format>
    <premis:formatDesignation>
      <premis:formatName>Extensible Hypertext Markup Language</premis:formatName>
      <premis:formatVersion>1</premis:formatVersion> <premis:formatRegistry>
        <premis:formatRegistryName>PRONOM</premis:formatRegistryName>
        <premis:formatRegistryKey>fmt/102</premis:formatRegistryKey>
        <premis:formatRegistryRole>specification</premis:formatRegistryRole>
      </premis:formatRegistry>
    </premis:format>
  </premis:objectCharacteristics>

```

## A.2. Ereignisbeispiel

Dieses Beispiel beschreibt ein Ereignis, das von einem Langzeitarchiv an einem Objekt durchgeführt wurde. Eine gängige Funktion eines Langzeitarchivs ist die Identifizierung des Dateiformats eines Objekts und die Validierung, dass es mit den Formatstandards übereinstimmt, wobei Übereinstimmung oder Ausnahmen festgestellt werden. Das Beispiel veranschaulicht die Formatvalidierung von Datei 1 im vorherigen Objektbeispiel.

| PREMIS Semantic Unit           | Datei 1  |
|--------------------------------|--|
| 2.1 eventIdentifier            |  |
| 2.1.1 eventIdentifierType      | DPS  |
| 2.1.1 eventIdentifierValue     | 25   |
| 2.2 eventType                  | validation   |
| 2.3 eventDateTime              | 2013-07-01T17:23:25Z   |
| 2.4 eventDetailInformation     |  |
| 2.4.1 eventDetail              | Format identification performed on file  |
| 2.5 eventOutcomeInformation    |  |
| 2.5.1 eventOutcome             | success  |
| 2.5.2 eventOutcomeDetail       |  |
| 2.5.2.1 eventOutcomeDetailNote | FORMAT_ID=fmt/102;IDENTIFICATION_METHOD=SIGNATURE;FILE_EXTENSION=html;DEPOSIT_ |

|                                    |  |
|------------------------------------|--|
|                                    | ACTIVITY_ID=659202;PID=FL26274026;SIP_ID=444381;PRODUCER_ID=29408010705;TASK_ID=48;PROCESS_ID=;MF_ID=8 |
| 2.6 linkingAgentIdentifier         |  |
| 2.6.1 linkingAgentIdentifierType   | local  |
| 2.6.2 linkingAgentIdentifierValue  | DROID6_63_1  |
| 2.6.3 linkingAgentIdentifierRole   | implementer  |
| 2.7 linkingObjectIdentifier        |  |
| 2.7.1 linkingObjectIdentifierType  | FILEPID  |
| 2.7.2 linkingObjectIdentifierValue | 17415493   |

Ein XML-Ausschnitt für das Ereignis der Formatidentifizierung könnte so aussehen:

```

<premis:event>
  <premis:eventIdentifier>
    <premis:eventIdentifierType>DPS</premis:eventIdentifierType>
    <premis:eventIdentifierValue>25</premis:eventIdentifierValue>
  </premis:eventIdentifier>
  <premis:eventType authority="event type"
    authorityURI="http://id.loc.gov/vocabulary/preservation/eventType"
    valueURI="http://id.loc.gov/vocabulary/preservation/eventType/val">
    validation</premis:eventType>
  <premis:eventDateTime>2013-07-01T17:23:25Z</premis:eventDateTime>
  <premis:eventDetailInformation>
    <premis:eventDetail>Format identification performed on file</premis:eventDetail>
  </premis:eventDetailInformation>
  <premis:eventOutcomeInformation>
    <premis:eventOutcome>success</premis:eventOutcome>
    <premis:eventOutcomeDetail> <premis:eventOutcomeDetailNote>
      FORMAT_ID=fmt/102;IDENTIFICATION_METHOD=SIGNATURE;FILE_EXTENSION=html;
      DEPOSIT_ACTIVITY_ID=659202;PID=FL26274026;SIP_ID=444381;
      PRODUCER_ID=29408010705;TASK_ID=48;PROCESS_ID=;MF_ID=8</premis:eventOutcomeDetailNote>
    </premis:eventOutcomeDetail>
  </premis:eventOutcomeInformation>
  <premis:linkingAgentIdentifier>
    <premis:linkingAgentIdentifierType>LOCAL</premis:linkingAgentIdentifierType>
    <premis:linkingAgentIdentifierValue>DROID6_63_1</premis:linkingAgentIdentifierValue>
    <premis:linkingAgentRole authority="eventRelatedAgentRole"
      authorityURI="http://id.loc.gov/vocabulary/preservation/eventRelatedAgentRole"
      valueURI="http://id.loc.gov/vocabulary/preservation/eventRelatedAgentRole/imp">
      implementer</premis:linkingAgentRole>
  </premis:linkingAgentIdentifier>
  <premis:linkingObjectIdentifier>
    <premis:linkingObjectIdentifierType>FILEPID</premis:linkingObjectIdentifierType>
    <premis:linkingObjectIdentifierValue>17415493</premis:linkingObjectIdentifierValue>
  </premis:linkingObjectIdentifier>
</premis:event>

```

### A.3. Agentenbeispiel

Dieses Beispiel beschreibt den Agenten, der mit dem oben beschriebenen Ereignis assoziiert ist. In diesem Fall ist der Agent eine Software, die die Aktion am Objekt ausgeführt hat.

| PREMIS Semantic Unit       |  |
|----------------------------|--|
| 3.1 agentIdentifier        |  |
| 3.1.1 agentIdentifierType  | local  |
| 2.1.1 agentIdentifierValue | DROID6_63_1  |
| 3.2 agentName              | REG_SA_DROID                                       |
| 3.3 agentType              | software   |
| 3.4 agentVersion           | Version 6.01                                       |
| 3.5 agentNote              | Signature version Binary SF v.63/ Container SF v.1 |

Ein XML-Ausschnitt für die Beschreibung eines Agenten:

```
<premis:agent>
  <premis:agentIdentifier>
    <premis:agentIdentifierType>local</premis:agentIdentifierType>
    <premis:agentIdentifierValue>DROID6_63_1</premis:agentIdentifierValue>
  </premis:agentIdentifier>
  <premis:agentName>REG_SA_DROID</premis:agentName>
  <premis:agentType>software</premis:agentType>
  <premis:agentVersion>Version 6.01</premis:agentVersion>
  <premis:agentNote>Signature version Binary SF v.63/ Container SF v.1</premis:agentNote> </premis:agent>
```

## Appendix B: 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.

**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.