

KOST.07

Archivierung von Unterlagen aus Geschäftsverwaltungssystemen (AUGev)

Vergleich LMER-PREMIS

Inhalt

1	Vorbemerkungen	1
2	Technische Metadaten.....	1
3	Verzeichnung von Migrationen	3

1 Vorbemerkungen

In der Referenzimplementierung der KOST wird für die administrativ-technischen Metadaten der internationale PREMIS-Standard verwendet. Das DIAS-System, welches den Kern des St. Galler Prototyps bildet, verwendet seinerseits den von der Deutschen Nationalbibliothek entwickelten LMER-Standard. Das vorliegende Dokument vergleicht die beiden Standards im Hinblick auf die Anforderungen, die sie im KOST-Prototypen abdecken sollen. Der Vergleich ist beschränkt und knapp gehalten. Ein ausführlicher Vergleich inkl. Mapping der beiden Standards ist von Olaf Brandt von der Königlichen Bibliothek in Den Haag unternommen worden und soll in Bälde von der nestor-AG Standards¹ publiziert werden.

2 Technische Metadaten

In der KOST-Referenzimplementierung werden die folgenden technischen Metadaten verzeichnet:

- URL/Dateiname
- Hashwert
- Dateigrösse
- MIME-Type
- Formatbezeichnung
- Entstehungsdatum (Creation Date)

Dazu muss angemerkt werden, dass praktisch alle diese Angaben nicht zwingend in PREMIS bzw. LMER gespeichert werden müssen: Für den Pfad zur Datei (URL/Dateiname), den Hashwert, die Dateigrösse und das Entstehungsdatum existieren beispielsweise eigene Metadatenfelder in METS. Zur Zeit ist nur für die Formatangaben (MIME-Type, Formatbezeichnung) ein Preservation-Metadatenchema notwendig, da es nicht möglich ist, diese sinnvoll in einem anderen Schema unterzubringen. (Siehe dazu im Detail die Beurteilungen verschiedener AIP-Lösungen.)

¹ http://nestor.cms.hu-berlin.de/moinwiki//AG_Langzeitarchivierungsstandards.

Die Verzeichnung dieser Angaben in PREMIS sieht an einem Beispiel folgendermaßen aus:

```
<object xmlID="i__97690472_519">
  <objectIdentifier>
    <objectIdentifierType>URL</objectIdentifierType>
    <objectIdentifierValue>{1572898D-1A79-4EF6-BA28-
881602B1780A}.pdf</objectIdentifierValue>
  </objectIdentifier>
  <objectCategory>file</objectCategory>
  <objectCharacteristics>
    <fixity>
      <messageDigestAlgorithm>MD5</messageDigestAlgorithm>
      <messageDigest>8c3041134fb16ad587e4d5f11bf8ac89</messageDigest>
    </fixity>
    <size>1999</size>
    <format>
      <formatDesignation>
        <formatName>application/pdf</formatName>
        <formatVersion>PDF document, version 1.3</formatVersion>
      </formatDesignation>
    </format>
  </objectCharacteristics>
  <creatingApplication>
    <dateCreatedByApplication>2007-03-
26T16:13:30</dateCreatedByApplication>
  </creatingApplication>
</object>
```

Das gleiche Beispiel in LMER verzeichnet:

```
<object ID="i__97734664_294">
  <lmerObject:objectIdentifier>ABC:AP
Räte:310:22.06.02</lmerObject:objectIdentifier>
  <lmerObject:numberOfFiles>1</lmerObject:numberOfFiles>
  <file>
    <lmerFile:fileIdentifier>{1572898D-1A79-4EF6-BA28-
881602B1780A}.pdf</lmerFile:fileIdentifier>
    <lmerFile:name>{1572898D-1A79-4EF6-BA28-
881602B1780A}.pdf</lmerFile:name>
    <lmerFile:size>1999</lmerFile:size>
    <lmerFile:fileDateTime>2006-04-02T00:00:00</lmerFile:fileDateTime>
    <lmerFile:fileChecksum CHECKSUM-
TYPE="MD5">8c3041134fb16ad587e4d5f11bf8ac89</lmerFile:fileChecksum>
    <lmerFile:mimeType>application/pdf</lmerFile:mimeType>
    <lmerFile:format REGISTRYNAME="PRONOM">fmt/17</lmerFile:format>
    <lmerFile:formatInfos>PDF document, version 1.3</lmerFile:formatInfos>
    <lmerFile:category>binary</lmerFile:category>
  </file>
</object>
```

Der Vergleich zeigt folgendes:

- Die aus unserer Sicht notwendigen technischen Metadaten lassen sich sowohl in PREMIS als auch in LMER abbilden.
- Das LMER-File enthält zwingend die aus Sicht der KOST-Referenzimplementierung nicht notwendigen, aber vom Schema verlangten Elemente `<numberOfFiles>` und `<category>`.
- Wesentliche Unterschiede sind nicht erkennbar. Bemerkenswert sind jedoch einige kleinere Unterschiede:
 - o Für die Darstellung des Formats werden verschiedene Ansätze gewählt. PREMIS stellt zwei Varianten zur Verfügung und verlangt, dass mindestens eine davon gewählt wird: die `<formatDesignation>`, eine Freitext-Beschreibung, für welche wir uns allerdings auf MIME-Typen beschränken, oder die `<formatRegistry>`, welche auf eine anerkannte Formatregistry (z.B. PRONOM²) verweist und den dortigen Key für das Format angibt. LMER kennt drei Elemente, die für die Formatbeschreibung verwendet werden können, allesamt fakultativ: `<mimeType>` für den MIME-Type; `<format>` für den Verweis auf eine Formatregistry, deren Name als Attribut mitgeliefert wird; und `<formatInfos>` als Freitextfeld, das im obigen Beispiel für Formatbezeichnung und Versionsnummer verwendet wird.
 - o Das Datum der Datei wird bei LMER umstandslos als `<fileDateTime>` verzeichnet. Bei PREMIS ist dafür das Feld `<dateCreatedByApplication>` innerhalb von `<creatingApplication>` vorgesehen.

3 Verzeichnung von Migrationen

Die korrekte Verzeichnung von Migrationen wurde im Projekt AUGev nicht prioritär behandelt. In Anbetracht der Tatsache, dass heute archivierte Unterlagen voraussichtlich nicht früher als in fünf bis zehn Jahren migriert werden müssen, wurde eine vertiefte Beschäftigung mit und Vorbereitung auf die Migration als überflüssig erachtet. Gleichwohl wurden einige Überlegungen angestellt. Deren Grundlage war, dass die Migration einer Primärdatei aus archivischen Gründen keine Veränderung im Findmittel, d.h. in den deskriptiven Metadaten, nach sich ziehen sollte. Es hat sich allerdings gezeigt, dass es nicht trivial ist, die Migrationen allein mittels *Preservation Metadata* zu verwalten.

In PREMIS werden aufeinanderfolgende Migrationen in den Einträgen zu den entsprechenden Dateien mit dem `<relationship>`-Element verknüpft. Eine Vorgängerdatei verweist dabei auf ihre Migration; eine migrierte Datei verweist auf ihren Vorgänger zurück. Aus dem EAD-File wird auf den PREMIS-Eintrag der ursprünglichen Instanz verwiesen. Um die aktuellste Migration zu identifizieren, springt man so lange zur nächsten Migration weiter, bis eine Datei keinen „wurde migriert in“-Eintrag in `<relationshipSubType>` mehr hat. Nähere Informationen zu einem Migrationsvorgang werden als `<event>` festgehalten.

```
<relationship>
  <relationshipType>Derivation</relationshipType>
  <relationshipSubType>wurde migriert in</relationshipSubType>
```

² <http://nationalarchives.gov.uk/pronom>.

```

    <relatedObjectIdentification>
      <relatedObjectIdentifierType>URL</relatedObjectIdentifierType>
      <relatedObjectIdentifierValue>{1572898D-1A79-4EF6-BA28-
881602B1780A}.pdf</relatedObjectIdentifierValue>
      <relatedObjectSequence>0</relatedObjectSequence>
    </relatedObjectIdentification>
    <relatedEventIdentification>
      <relatedEventIdentifierType>Eventnummer</relatedEventIdentifierType>
      <relatedEventIdentifierValue>0</relatedEventIdentifierValue>
      <relatedEventSequence>0</relatedEventSequence>
    </relatedEventIdentification>
  </relationship>

<event>
  <eventIdentifier>
    <eventIdentifierType>Eventnummer</eventIdentifierType>
    <eventIdentifierValue>0</eventIdentifierValue>
  </eventIdentifier>
  <eventType>Migration</eventType>
  <eventDateTime>2008-01-16T17:13:23</eventDateTime>
</event>

```

In LMER entspricht der Abschnitt "Prozess" in etwa dem Event:

```

<process>
  <lmerProcess:oldObjectIdentifier>{1572898D-1A79-4EF6-BA28-
881602B1780A}.pdf</lmerProcess:oldObjectIdentifier>
  <lmerProcess:purpose>PDF 1.3 wird nicht mehr als archivtaugliches Format
akzeptiert.</lmerProcess:purpose>
  <lmerProcess:processCreator>PDF/A-Maker</lmerProcess:processCreator>
  <lmerProcess:steps>1. Analyse und Validierung des vorliegenden Formats,
2. Transformation in PDF/A, 3. Validierung des PDF/A-Files</lmerProcess:steps>
  <lmerProcess:result>File liegt im Format PDF/A vor.</lmerProcess:result>
  <lmerProcess:completionDate>2008-04-
02T13:45:07</lmerProcess:completionDate>
</process>

```

Problematisch ist hier einerseits die Tatsache, dass die meisten dieser Felder Freitextfelder sind, andererseits der Verweis von der originalen auf die migrierte Datei. Möglich ist nur der Rückverweis von der migrierten auf die originale Datei. Dies macht eine unerwünschte Anpassung des Findmitteleintrags notwendig.