The Matterhorn RDF Data Model:
Implementing OAIS and RiC in the context of semantic technologies

Alain Dubois, Andreas Nef
10 years and counting, two streams of reflection

• At the core, the concept of digital preservation, which – particularly thanks to contextualization – ensures the quality of information over its entire life cycle
• The principle of complying with international norms and standards
Preserving data

- OAIS reference model
Contextualized archival description

- Description standards of ICA
Matterhorn METS Profile

• First actual application of the conceptual models of preservation and contextualization
• A toolkit consisting of
  – a Metadata-Map
  – the Matterhorn METS Profile
  – open source tools (docuteam packer, docuteam feeder)
Matterhorn METS Profile: Metadata-Map
Matterhorn METS Profile

- Registered and published in 2012 by the Library of Congress as a generic profile that can be used by memory institutions (LAMs) for the management of archival fonds

- Open source tools (docuteam packer, docuteam feeder), used in more than 20 archival institutions

- Use of XML for serialization
Improvements and developments

- Rework the Matterhorn METS Profile in view of the possibilities offered by Linked Data and Open Linked Data, for which digital preservation is a core motivation
- A larger context (libraries, Wikipedia etc.) for archives
- Make use of existing resources to improve the precision of archival description
New technologies

- Lack of models that provide links between various ICA-standards
- XML is not sufficient to link these resources
- Semantic technologies are becoming the standard for archival description and library catalogues
Reflections by EGAD since 2012

- More than a standard for archival description
- Aligned to the concept of *Records in Context*
A changing world that entails an overhaul

• The Matterhorn METS profile needs to be revised in view of the body of work originating from EGAD and the archival community and of the constant evolution of technology

• Choosing usable standards for long term preservation, independent from the institution that applies them (archives, libraries, museums, research centres, documentation centres)
Two different approaches

- **EGAD’s nongeneric approach**: Developing an RDF standard specific to archives, with gateways to library and museum standards.

- **The Matterhorn RDF Data Model’s generic approach**: Based upon RDF existing and consensual international standards, allowing to model Records in Context.

- In contrast to EGAD, the Matterhorn RDF Data Model is based on existing ontologies. It follows the best practices propagated by the W3C: «It is best practice to use or extend an existing vocabulary before creating a new vocabulary.»
2017 – Matterhorn RDF Data Model
Matterhorn RDF Data Model

RiC-E1 Record
RiC-E2 Record Component
RiC-E3 Record Set

Object, Entity (technical)
premis:Object

Object, Entity (description)
pedm:Object

Rights
premis:RightsStatement

File
pedm:File

Agent (technical)
premis:Agent

Agent (description)
rda:C10002

Event
premis:Event

Activity, Function
bpmn:Activity

RiC-E4 Agent
RiC-E5 Occupation
RiC-E6 Position

RiC-E28 History

RiC-E7 Function
RiC-E8 Function (Abstract)
RiC-E9 Activity
RiC-E10 Mandate

Technical Metadata
Descriptive Metadata
Matterhorn RDF Data Model

### Classes

<table>
<thead>
<tr>
<th>Intellectual Concept</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity</td>
<td>bpmn:Activity</td>
</tr>
<tr>
<td>Agent</td>
<td>premis:Agent, rdac:C10002</td>
</tr>
<tr>
<td>Event</td>
<td>premis:Event</td>
</tr>
<tr>
<td>File</td>
<td>premis:Object, pcdm:File</td>
</tr>
<tr>
<td>Identifier</td>
<td>premis:Identifier</td>
</tr>
<tr>
<td>Object, Entity, Record</td>
<td>premis:Object, pcdm:Object</td>
</tr>
<tr>
<td>Physical Location</td>
<td>premis:ContentLocation</td>
</tr>
<tr>
<td>Rights</td>
<td>premis:RightsStatement</td>
</tr>
</tbody>
</table>
An important portion of the properties used in the Matterhorn RDF Data Model originate from the RDA (Resource Description and Access) standard.
## Object, Entity (description)

**Class:** pcdm:Object

### ISAD(G)

<table>
<thead>
<tr>
<th>Property</th>
<th>Values</th>
<th>Standard</th>
<th>Accessor</th>
<th>EAD 2002</th>
</tr>
</thead>
<tbody>
<tr>
<td>dc:identifier</td>
<td>Literal</td>
<td>ISAD 1.1</td>
<td>refCode</td>
<td>EAD:did/EAD:unitid[@type = 'refCode']</td>
</tr>
<tr>
<td>dc:title</td>
<td>Literal</td>
<td>ISAD 1.2</td>
<td>unitTitle</td>
<td>EAD:did/EAD:unittitle[@label = 'main']</td>
</tr>
<tr>
<td>dc:date</td>
<td>Literal</td>
<td>ISAD 1.3</td>
<td>date</td>
<td>EAD:did/EAD:unitdate[@label = 'date']</td>
</tr>
<tr>
<td>dct:temporal</td>
<td>Literal</td>
<td>ISAD 1.3 Betrifftzeitraum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dc:type</td>
<td>Literal</td>
<td>ISAD 1.4</td>
<td>otherLevelName</td>
<td></td>
</tr>
<tr>
<td>dcterms:isPartOf</td>
<td>Literal</td>
<td>ISAD 1.5</td>
<td>extent</td>
<td></td>
</tr>
<tr>
<td>dct:provenance</td>
<td>Literal or URI of rdac:C10002</td>
<td>ISAD 2.1</td>
<td>origination</td>
<td></td>
</tr>
<tr>
<td>rdau:P60484</td>
<td>Literal</td>
<td>ISAD 2.2; RDA „has agent history“</td>
<td>biographicalHistory</td>
<td>EAD:bioghist/EAD:p</td>
</tr>
<tr>
<td>rdau:P60176</td>
<td>Literal</td>
<td>ISAD 2.3; RDA „has custodial history of resource“</td>
<td>archivalHistory</td>
<td>EAD:custodhist/EAD:p</td>
</tr>
<tr>
<td>rdau:P60583</td>
<td>Literal</td>
<td>ISAD 2.4; RDA „has immediate source of acquisition of resource“</td>
<td>accessNr (eigentlich acquisitionInfo)</td>
<td>EAD:acqinfo/EAD:p</td>
</tr>
<tr>
<td>dc:description</td>
<td>Literal</td>
<td>ISAD 3.1</td>
<td>scopeContent</td>
<td>EAD:scopecontent/EAD:p</td>
</tr>
<tr>
<td>dcterms:isPartOf</td>
<td>Literal</td>
<td>ISAD 3.2</td>
<td>appraisalAndDestruction</td>
<td>EAD:appraisal/EAD:p</td>
</tr>
<tr>
<td>dcterms:isPartOf</td>
<td>Literal</td>
<td>ISAD 3.3</td>
<td>accruals</td>
<td>EAD:accruals/EAD:p</td>
</tr>
<tr>
<td>rdau:P60348</td>
<td>Literal</td>
<td>ISAD 3.4; RDA „has system of arrangement“</td>
<td>arrangement</td>
<td>EAD:arrangement/EAD:p</td>
</tr>
</tbody>
</table>
Example 1: Inventory database to descriptive metadata

We’re currently testing an automatic generation of triples based on archival descriptions managed in docuteam curator for export to aLOD (www.alod.ch). The content is the inventory of the town of Baden.

The process:

• The standard «RDB to RDF Mapping Language» (R2RML) is used for the transformation of the «raw» data recorded in the database from one model to the other.

• Zazuko, the company in charge of the aLOD project, recommends Turtle syntax for expressing data in RDF triples. The data is stored in a database provided by a commercial triple store called Stardog.
Table Archival Description (docuteam curator MySQL)

Archival descriptions structured according to ISAD(G). Other tables coming from the same database provide the link to ISAAR-CPF authority records.
Individual triples representing an object

URI of a series of “objects” after transformation to RDF
Triples representing ISAD(G) elements as resource properties


<table>
<thead>
<tr>
<th>p</th>
<th>o</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://purl.org/dc/terms/identifier">http://purl.org/dc/terms/identifier</a></td>
<td>68</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/title">http://purl.org/dc/terms/title</a></td>
<td>Zingg, Familie, Postkartenverlag</td>
</tr>
<tr>
<td><a href="http://purl.org/dc/terms/description">http://purl.org/dc/terms/description</a></td>
<td>Familiendokumente, Fotos, Postkarten</td>
</tr>
<tr>
<td>rdf:type</td>
<td><a href="http://pcdm.org/models#Object">http://pcdm.org/models#Object</a></td>
</tr>
<tr>
<td><a href="http://pcdm.org/models#hasRelatedObject">http://pcdm.org/models#hasRelatedObject</a></td>
<td><a href="http://data.alod.ch/stabaden/record/129">http://data.alod.ch/stabaden/record/129</a></td>
</tr>
</tbody>
</table>
Same principle for an archival description of authority records structured according to ISAAR-CPF

\[<http://data.alod.ch/stabaden/person/900> ?p ?o\]

<table>
<thead>
<tr>
<th>SPARQL Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>p</td>
</tr>
<tr>
<td><a href="http://rdaregistry.info/Elements/u/P60368">http://rdaregistry.info/Elements/u/P60368</a></td>
</tr>
<tr>
<td><a href="http://rdaregistry.info/Elements/u/P60598">http://rdaregistry.info/Elements/u/P60598</a></td>
</tr>
<tr>
<td><a href="http://rdaregistry.info/Elements/u/P60599">http://rdaregistry.info/Elements/u/P60599</a></td>
</tr>
<tr>
<td>rdf:type</td>
</tr>
</tbody>
</table>

RDA „has date of death“
RDA „has date of birth“
Exemple 2: Conversion Matterhorn METS → Matterhorn RDF (dossier administratif)
# ====== files (<METS:fileSec>)

<971491.pdf> a pcdm:File;
dcterms:extent "5857936";
dcterms:format "fmt/18".

<9718207.pdf> a pcdm:File;
dcterms:extent "2203813";
dcterms:format "fmt/18".

<971940.pdf> a pcdm:File;
dcterms:extent "1021785";
dcterms:format "fmt/18".

<9723001.pdf> a pcdm:File;
dcterms:extent "2203813";
dcterms:format "fmt/18".

<9723417.pdf> a pcdm:File;
dcterms:extent "5857936";
dcterms:format "fmt/18".

<9718786.doc> a pcdm:File;
dcterms:extent "64512";
dcterms:format "fmt/48".

<9725379.pdf> a pcdm:File;
dcterms:extent "1236958";
dcterms:format "fmt/276".

# ====== agents

<20170907150515477> a premis:Agent;
  rdau:P60368 "Muster, Daniel";
  rdau:P60053 "musdan".

<20170902340515791> a premis:Agent;
  rdau:P60368 "Example, Peter";
  rdau:P60053 "exapet".

# ====== events (<PREMIS:event>)

<20170907150515415> a premis:Event;
  premis:hasEventRelatedObject <9710632>;
  premis:hasEventRelatedAgent <20170907150515477>;
  premis:hasEventType "01.01 - Préparation";
  premis:hasEventOutcomeInformation "COMPLETED".

<20170907150515633> a premis:Event;
  premis:hasEventRelatedObject <9710632>;
  premis:hasEventRelatedAgent <20170907150515477>;
  premis:hasEventType "02.01 - Analyse formelle du dossier";
  premis:hasEventOutcomeInformation "COMPLETED".

<20170907150515791> a premis:Event;
  premis:hasEventRelatedObject <9710632>;
  premis:hasEventRelatedAgent <20170907150515477>;
  premis:hasEventType "02.02 - Pré-Validation";
  premis:hasEventOutcomeInformation "COMPLETED".
And then?
Matterhorn RDF to RiC-O Crossover

To show that the RDF data model is considerably less rigid than, for instance, a relational database or serialized XML we want to create a mapping schema, which transforms semantic information stored as Matterhorn RDF according to a future Records in Context ontology.
Conclusion

• Digital preservation is a work in progress

• Our aim is to contribute to the ongoing discussions and reflections within the International Council on Archives

• We’re not interested in developing an alternative model

• We engage in professional debate with the members of EGAD
Contact information

Alain Dubois
Archiviste cantonal
Rue de Lausanne 45
1950 Sion
+41 27 606 46 05
alain.dubois@admin.vs.ch

Andreas Nef
Im Langacker 16
5405 Baden-Dättwil
+41 56 511 11 77
a.nef@docuteam.ch