

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

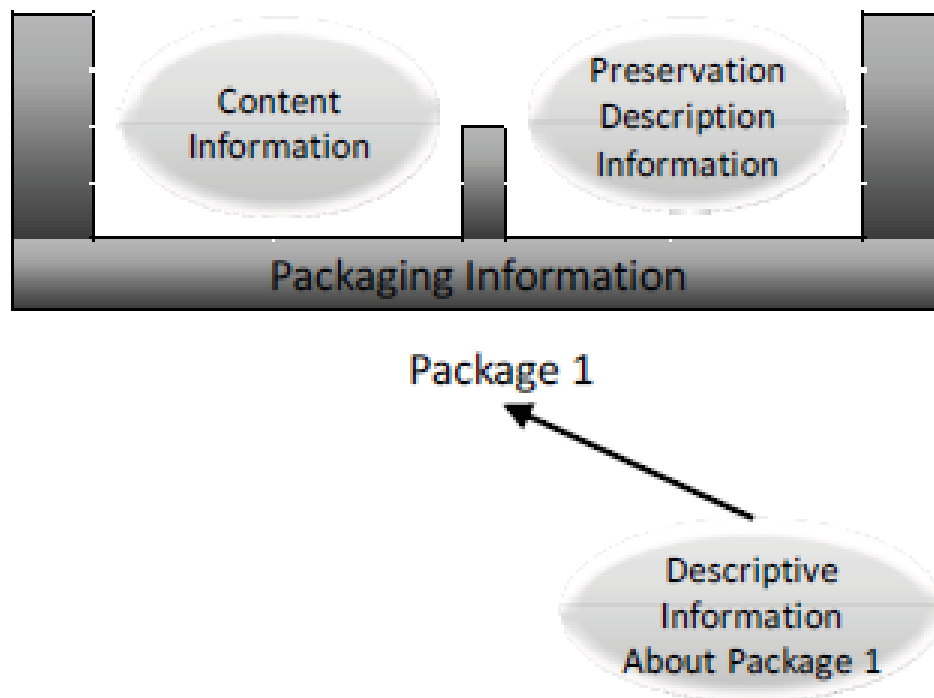
Alain Dubois, Andreas Nef

# 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

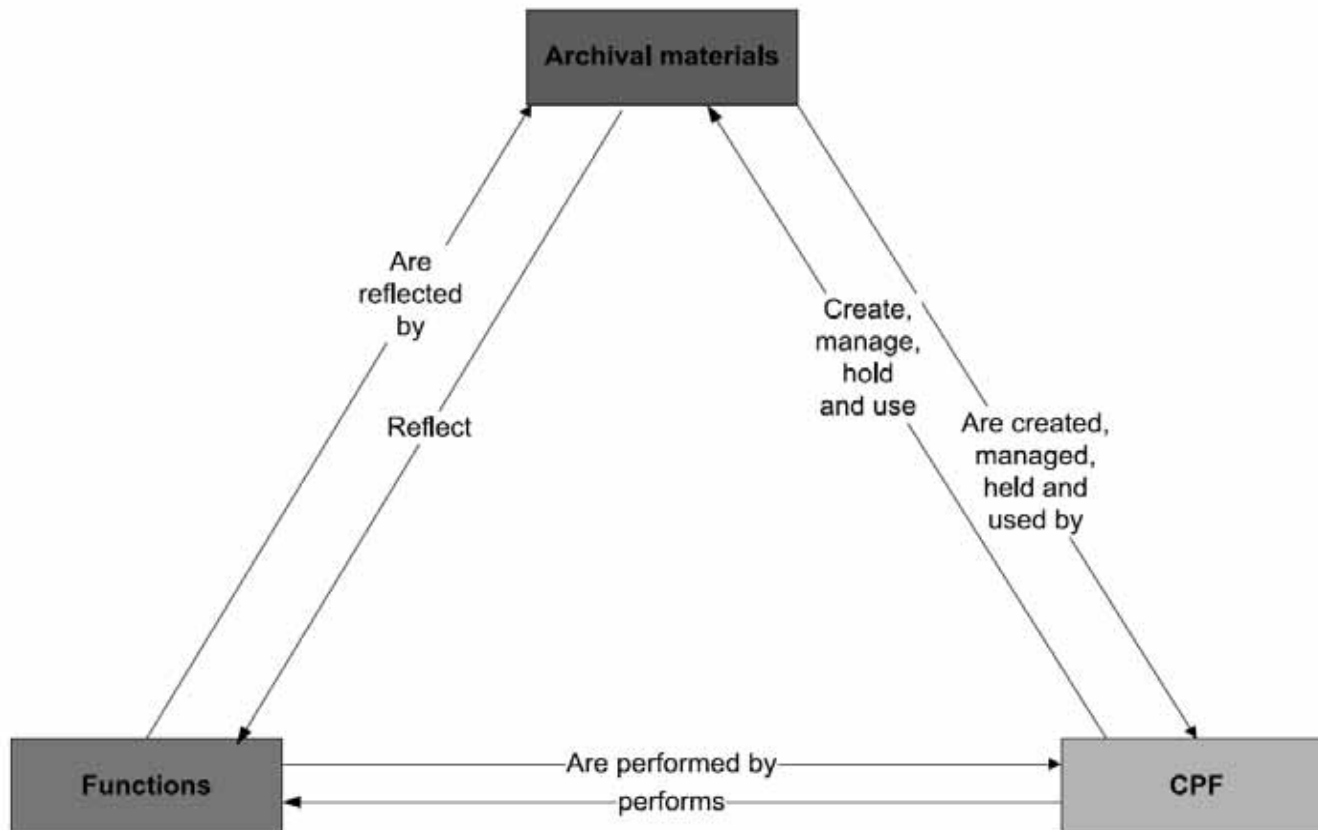
# Preserving data

- OAIS reference model



# Contextualized archival description

- Description standards of ICA





# 2012 – Matterhorn METS Profile

# Matterhorn METS Profile

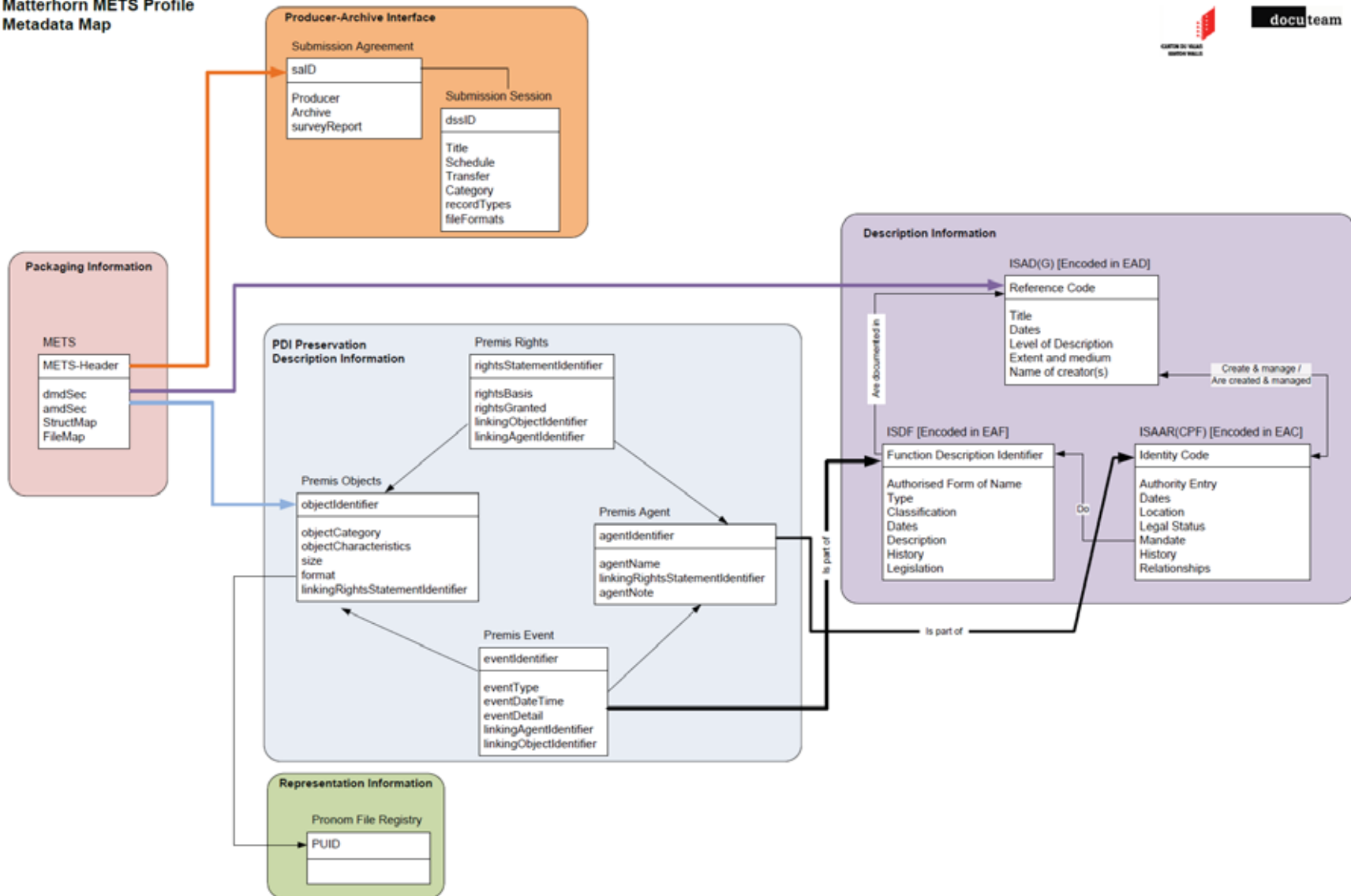
- First actual application of the conceptual models of preservation and contextualization
- A toolkit consisting of
  - a Metadata-Map
  - the Matterhorn METS Profile
  - open source tools (docuteam packer, docuteam feeder)

# Matterhorn METS Profile: Metadata-Map

Matterhorn METS Profile  
Metadata Map



docu team



# Matterhorn METS Profile

- 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)

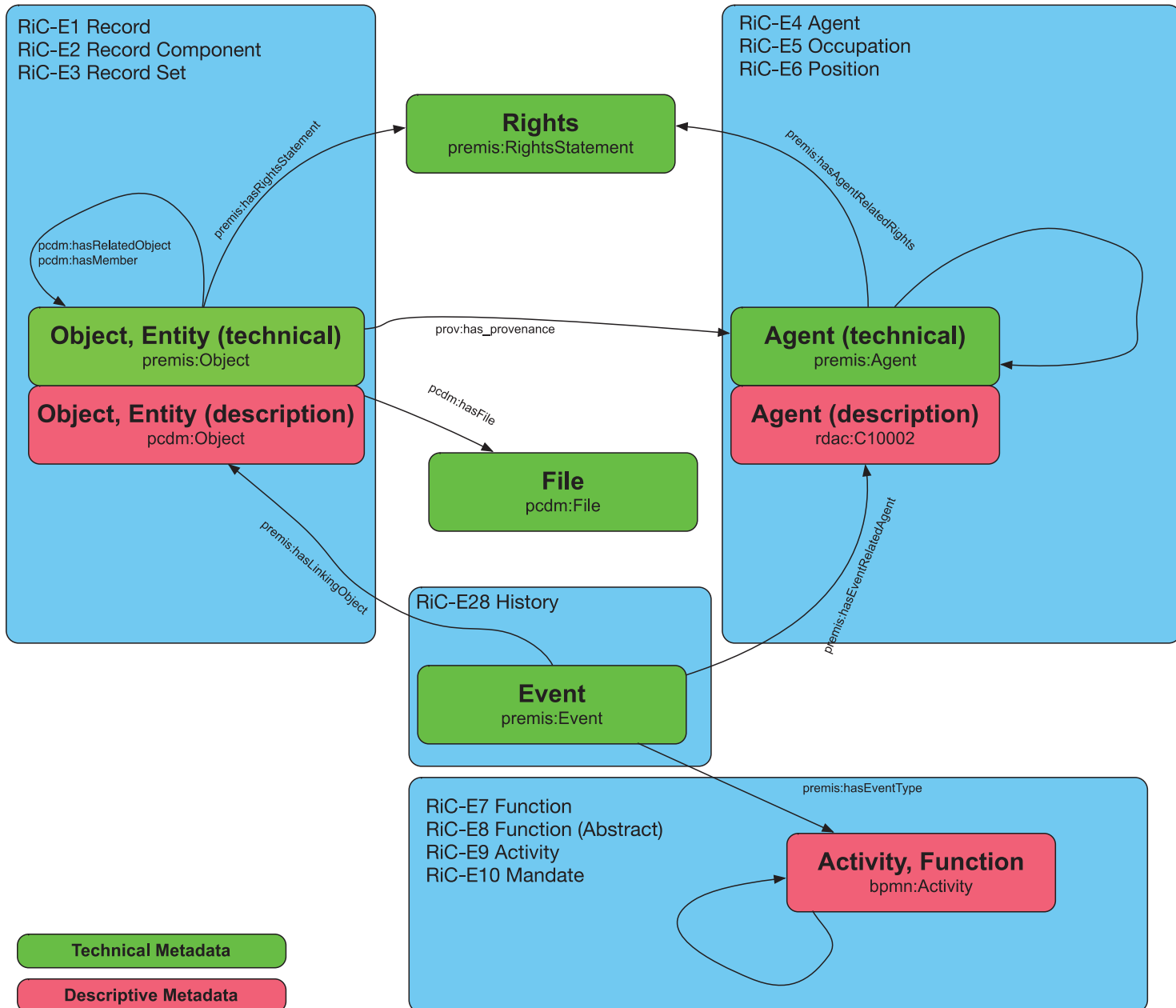
# Two different approaches

- **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.»*

A photograph of the Matterhorn mountain peak, a prominent, jagged, and snow-capped mountain peak in the Alps. The mountain is the central focus, with its sharp, rocky ridges and snow-covered slopes. The sky is a mix of blue and white, with large, billowing clouds partially obscuring the peak. The foreground shows dark, rocky terrain with some snow patches and a small stream or waterfall. The overall scene is dramatic and high-altitude.

# 2017 – Matterhorn RDF Data Model

# Matterhorn RDF Data Model



# Matterhorn RDF Data Model

## Classes

Intellectual Concept	Class
Activity	bpmn:Activity
Agent	premis:Agent, rdac:C10002
Event	premis:Event
File	premis:Object, pcdm:File
Identifier	premis:Identifier
Object, Entity, Record	premis:Object, pcdm:Object
Physical Location	premis:ContentLocation
Rights	premis:RightsStatement



# Matterhorn RDF Data Model

## Namespaces

Name	Prefix	URI
Business Process Model and Notation	bpmn	<a href="http://dkm.fbk.eu/index.php/BPMN_Ontology#">http://dkm.fbk.eu/index.php/BPMN_Ontology#</a>
Dublin Core	dc	<a href="http://purl.org/dc/elements/1.1/">http://purl.org/dc/elements/1.1/</a>
DC Terms	dct	<a href="http://purl.org/dc/terms">http://purl.org/dc/terms</a>
Portland Common Data Model	pcdm	<a href="http://pcdm.org/models#">http://pcdm.org/models#</a>
Premis	premis	<a href="http://www.loc.gov/premis/rdf/v1#">http://www.loc.gov/premis/rdf/v1#</a>
RDA Unconstrained properties	rdau	<a href="http://rdaregistry.info/Elements/u/">http://rdaregistry.info/Elements/u/</a>
RDA Class hierarchies	rdac	<a href="http://rdaregistry.info/Elements/c/">http://rdaregistry.info/Elements/c/</a>

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)

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

# 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](http://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.

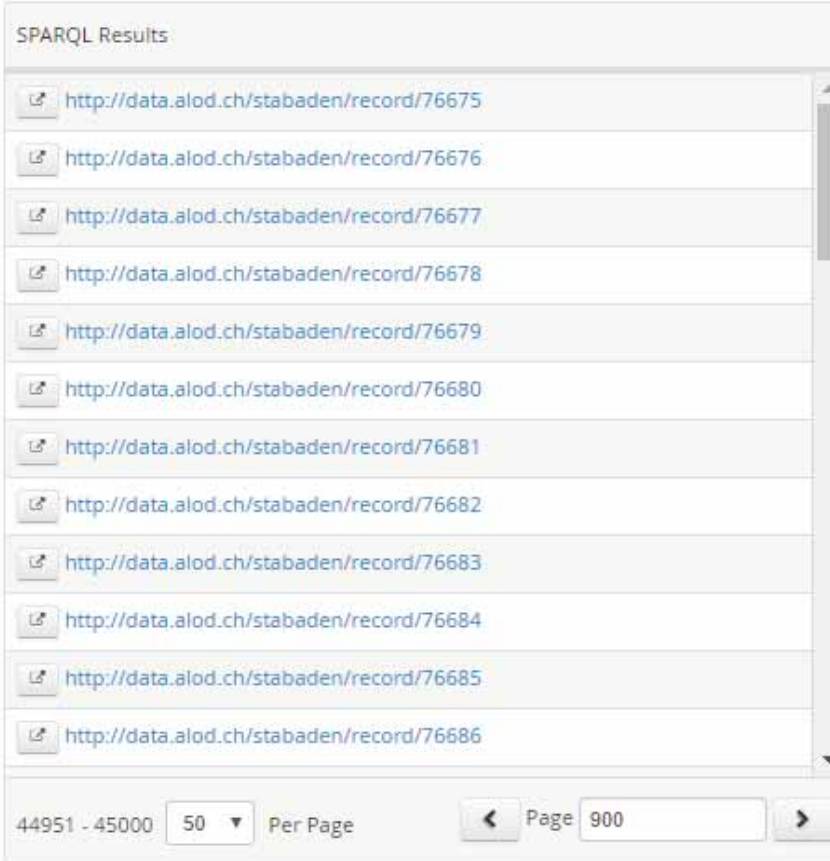
# Table Archival Description (docuteam curator MySQL)

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

id *	parent_id	11_reference_code *	14_level_of_description_id *	12_title	13_date_from	13_date_to
6244	129	61		10 Spoerry, Albert, Fabrikant, 1845-1918	{null}	{null}
6245	129	62		10 Gläser, Louise, Schneiderin, 1904-1995	{null}	{null}
6246	129	63		10 Frey, Karl, von Zurzach	18280000	18300000
6247	129	64		10 Jeuch, Jean	18240000	18280000
6248	129	65		10 Herzog, Josef, Färber in Ennetbaden	18450000	18500000
6249	129	66		10 Hafter, Albert, Oberingenieur BBC	19170000	19340000
6250	129	67		10 Wyler, Josef, Veit und Berthold	{null}	{null}
6251	129	68		10 Zingg, Familie, Postkartenverlag	{null}	{null}
6252	129	69		10 Surläuli, Mauriz gegen Familie Josef S...	18330000	18340000
6253	129	70		10 Tschudi auf Schwarzwasserstelz, Fam...	{null}	{null}
6254	129	71		10 Räber, Albert, Stadtrat	19860000	{null}
6255	129	72		10 Raschle, Hans, Stadtschreiber, Briefe ...	19130000	19160000
6256	129	73		10 Weibel, Werner, Schützen- und Turn...	19200000	19800000
6257	129	74		10 Pfyffer, Ivo, Unterlagen Seminar Wett...	18360000	19200000
6258	129	75		10 Steiner, Joseph, Werkmeister, techn. ...	19300000	19400000
6259	129	76		10 Zollinger, Edwin, Fotoalbum Bahnelek...	19050000	{null}
6260	129	77		10 Zollinger, Edwin, Fotoalbum Bahnelek...	{null}	{null}
6261	129	78		10 Zollinger, Edwin, Fotoalbum Grenzbe...	19140000	19180000
6262	129	79		10 Zollinger, Edwin, Album mit Postkarte...	{null}	{null}

# Individual triples representing an object

URI of a series of “objects” after transformation to RDF



SPARQL Results

<a href="http://data.alod.ch/stabaden/record/76675">http://data.alod.ch/stabaden/record/76675</a>
<a href="http://data.alod.ch/stabaden/record/76676">http://data.alod.ch/stabaden/record/76676</a>
<a href="http://data.alod.ch/stabaden/record/76677">http://data.alod.ch/stabaden/record/76677</a>
<a href="http://data.alod.ch/stabaden/record/76678">http://data.alod.ch/stabaden/record/76678</a>
<a href="http://data.alod.ch/stabaden/record/76679">http://data.alod.ch/stabaden/record/76679</a>
<a href="http://data.alod.ch/stabaden/record/76680">http://data.alod.ch/stabaden/record/76680</a>
<a href="http://data.alod.ch/stabaden/record/76681">http://data.alod.ch/stabaden/record/76681</a>
<a href="http://data.alod.ch/stabaden/record/76682">http://data.alod.ch/stabaden/record/76682</a>
<a href="http://data.alod.ch/stabaden/record/76683">http://data.alod.ch/stabaden/record/76683</a>
<a href="http://data.alod.ch/stabaden/record/76684">http://data.alod.ch/stabaden/record/76684</a>
<a href="http://data.alod.ch/stabaden/record/76685">http://data.alod.ch/stabaden/record/76685</a>
<a href="http://data.alod.ch/stabaden/record/76686">http://data.alod.ch/stabaden/record/76686</a>

44951 - 45000  Per Page  Page

# Triples representing ISAD(G) elements as resource properties

<<http://data.alod.ch/stabaden/record/6251>> ?p ?o

SPARQL Results	
p	o
<a href="http://purl.org/dc/terms/provenance">http://purl.org/dc/terms/provenance</a>	<a href="http://data.alod.ch/stabaden/person/28">http://data.alod.ch/stabaden/person/28</a>
<a href="http://purl.org/dc/terms/identifier">http://purl.org/dc/terms/identifier</a>	68
<a href="http://purl.org/dc/terms/title">http://purl.org/dc/terms/title</a>	Zingg, Familie, Postkartenverlag
<a href="http://purl.org/dc/terms/description">http://purl.org/dc/terms/description</a>	Familiendokumente, Fotos, Postkarten
<a href="http://rdaregistry.info/Elements/u/P60333">http://rdaregistry.info/Elements/u/P60333</a>	Doppler, Hugo; Zander, Edi. Baden: Seine weite Gasse. Baden 1987. Scherer, Walter; Füllemann, Verena. Baden um die Jahrhundertwende. Baden 1979. Badener Neujahrsblätter, 1978.
<a href="#">rdf:type</a>	<a href="http://pcdm.org/models#Object">http://pcdm.org/models#Object</a>
<a href="http://pcdm.org/models#hasRelatedObject">http://pcdm.org/models#hasRelatedObject</a>	<a href="http://data.alod.ch/stabaden/record/129">http://data.alod.ch/stabaden/record/129</a>

# Same principle for an archival description of authority records structured according to ISAAR-CPF

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

SPARQL Results	
p	o
<a href="http://rdaregistry.info/Elements/u/P60368">http://rdaregistry.info/Elements/u/P60368</a>	Squarise, Marco
<a href="http://rdaregistry.info/Elements/u/P60598">http://rdaregistry.info/Elements/u/P60598</a>	20090517
<a href="http://rdaregistry.info/Elements/u/P60599">http://rdaregistry.info/Elements/u/P60599</a>	19430000
<a href="http://rdaregistry.info/Elements/u/P60484">http://rdaregistry.info/Elements/u/P60484</a>	Künstler und Szenograf, Baden. Nachruf siehe Badener Neujaarsblätter 85 (2010), 210-212.
<a href="#">rdf:type</a>	<a href="http://rdaregistry.info/Elements/c/C10004">http://rdaregistry.info/Elements/c/C10004</a>


RDA „has date of death“


RDA „has date of birth“



# Exemple 2: Conversion Matterhorn METS → Matterhorn RDF (dossier administratif)

```
1 @prefix pcdm: <http://pcdm.org/models#>.
2 @prefix premis: <http://www.loc.gov/premis/rdf/v1#>.
3 @prefix dcterms: <http://purl.org/dc/terms/>.
4 @prefix dc: <http://purl.org/dc/elements/1.1/>.
5 @prefix rdac: <http://rdaregistry.info/Elements/c/>.
6 @prefix rdau: <http://rdaregistry.info/Elements/u/>.
7
8 # ===== mappings
9
10 pcdm:Object a premis:Object .
11 premis:Agent a rdac:C10002 .
12
13
14 # ===== structure (<METS:structMap>)
15
16 <9710632> a pcdm:Object ;
17   pcdm:hasMember <9710654> ;
18   pcdm:hasMember <9710656> ;
19   pcdm:hasMember <9725637> ;
20   pcdm:hasMember <9718786> ;
21   pcdm:hasMember <9725379> ;
22
23   dc:identifiant "2011.01290" ;
24   dc:title "Provisorische Ernennung Mr. X Z" ;
25   dc:description "Ernennung Mr. X, YYYY, wohnhaft in A als Z, Stelle Nr NNNNNN".
26
27 <9710654> a pcdm:Object ;
28   pcdm:hasMember <9717491> ;
29   pcdm:hasMember <9718207> .
30
31 <9717491> a pcdm:Object ;
32   pcdm:hasFile <9717491.pdf>.
33 <9718207> a pcdm:Object ;
34   pcdm:hasFile <9718207.pdf>.
35
```

```
55 # ===== files (<METS:fileSec>)
56
57 <9717491.pdf> a pcdm:File ;
58   dcterms:extent "5857936" ;
59   dc:format "fmt/18" .
60 <9718207.pdf> a pcdm:File ;
61   dcterms:extent "2203813" ;
62   dc:format "fmt/18" .
63 <9711940.pdf> a pcdm:File ;
64   dcterms:extent "1021785" ;
65   dc:format "fmt/18" .
66 <9723001.pdf> a pcdm:File ;
67   dcterms:extent "2203813" ;
68   dc:format "fmt/18" .
69 <9723417.pdf> a pcdm:File ;
70   dcterms:extent "5857936" ;
71   dc:format "fmt/18" .
72 <9718786.doc> a pcdm:File ;
73   dcterms:extent "64512" ;
74   dc:format "fmt/40" .
75 <9725379.pdf> a pcdm:File ;
76   dcterms:extent "1236950" ;
77   dc:format "fmt/276" .
78
```



```
55 # ===== files (<METS:fileSec>)
56
57 <9717491.pdf> a pcdm:File ;
58   dcterms:extent "5857936" ;
59   dc:format "fmt/18" .
60 <9718207.pdf> a pcdm:File ;
61   dcterms:extent "2203813" ;
62   dc:format "fmt/18" .
63 <9711940.pdf> a pcdm:File ;
64   dcterms:extent "1021785" ;
65   dc:format "fmt/18" .
66 <9723001.pdf> a pcdm:File ;
67   dcterms:extent "2203813" ;
68   dc:format "fmt/18" .
69 <9723417.pdf> a pcdm:File ;
70   dcterms:extent "5857936" ;
71   dc:format "fmt/18" .
72 <9718786.doc> a pcdm:File ;
73   dcterms:extent "64512" ;
74   dc:format "fmt/40" .
75 <9725379.pdf> a pcdm:File ;
76   dcterms:extent "1236950" ;
77   dc:format "fmt/276" .
78
```

```
80 # ===== agents
81
82 <20170907150515477> a premis:Agent ;
83   rdau:P60368 "Muster, Daniel" ;
84   rdau:P60053 "musdan" .
85
86 <20170902340515791> a premis:Agent ;
87   rdau:P60368 "Example, Peter" ;
88   rdau:P60053 "exapet" .
89
90 # ===== events (<PREMIS:event>)
91
92 <20170907150515415> a premis:Event ;
93   premis:hasEventRelatedObject <9710632> ;
94   premis:hasEventRelatedAgent <20170907150515477> ;
95   premis:hasEventType "01.01 - Préparation" ;
96   premis:hasEventOutcomeInformation "COMPLETED" .
97 <20170907150515633> a premis:Event ;
98   premis:hasEventRelatedObject <9710632> ;
99   premis:hasEventRelatedAgent <20170907150515477> ;
100  premis:hasEventType "02.01 - Analyse formelle du dossier" ;
101  premis:hasEventOutcomeInformation "COMPLETED" .
102 <20170907150515791> a premis:Event ;
103   premis:hasEventRelatedObject <9710632> ;
104   premis:hasEventRelatedAgent <20170907150515477> ;
105   premis:hasEventType "02.02 - Pré-Validation" ;
106   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