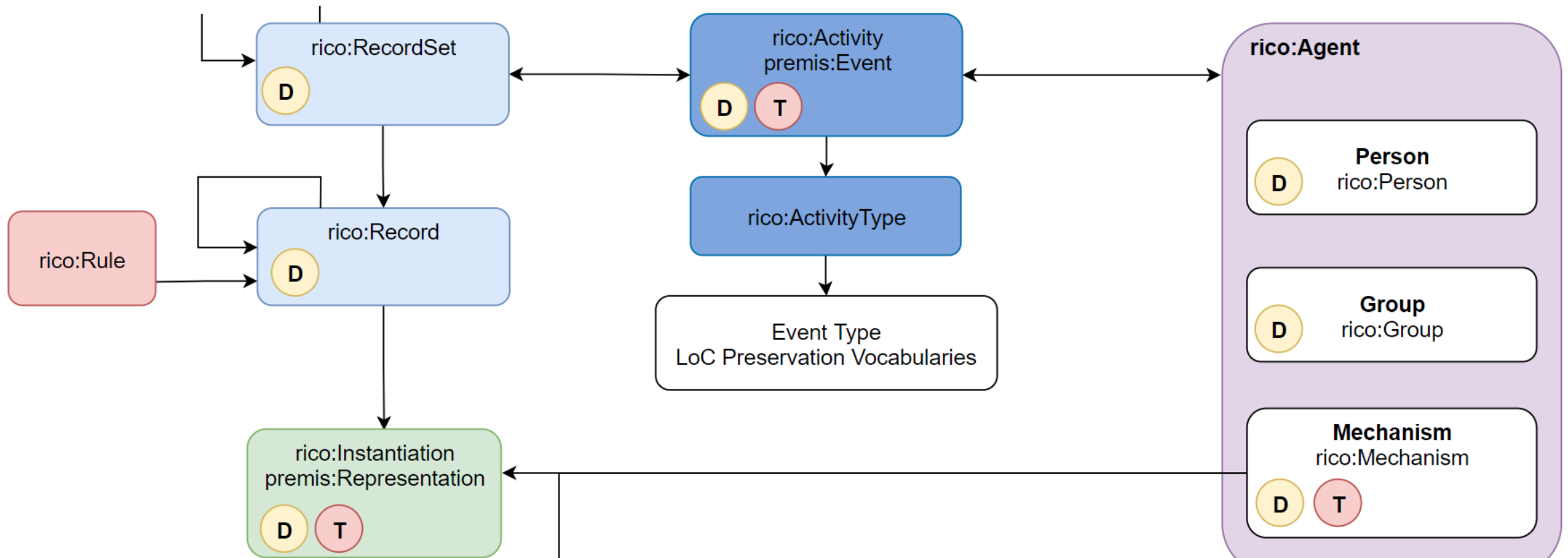


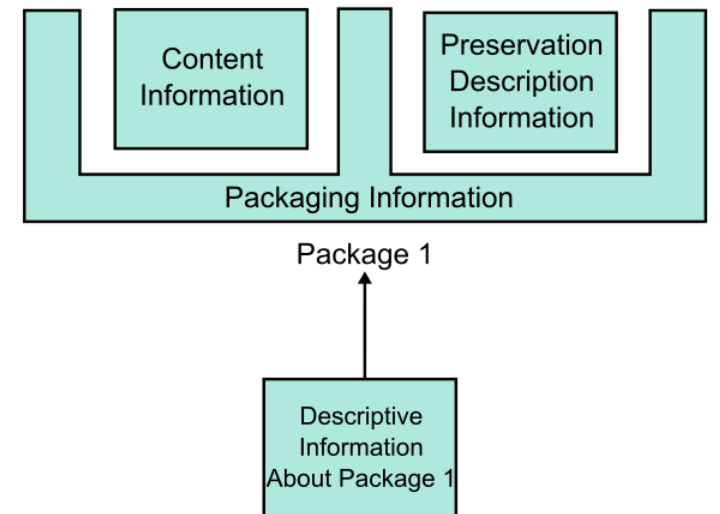
Matterhorn RDF: Contextualize archival records with RiC and PREMIS

ICA 2021 Virtual Conference – Empowering Knowledge Societies

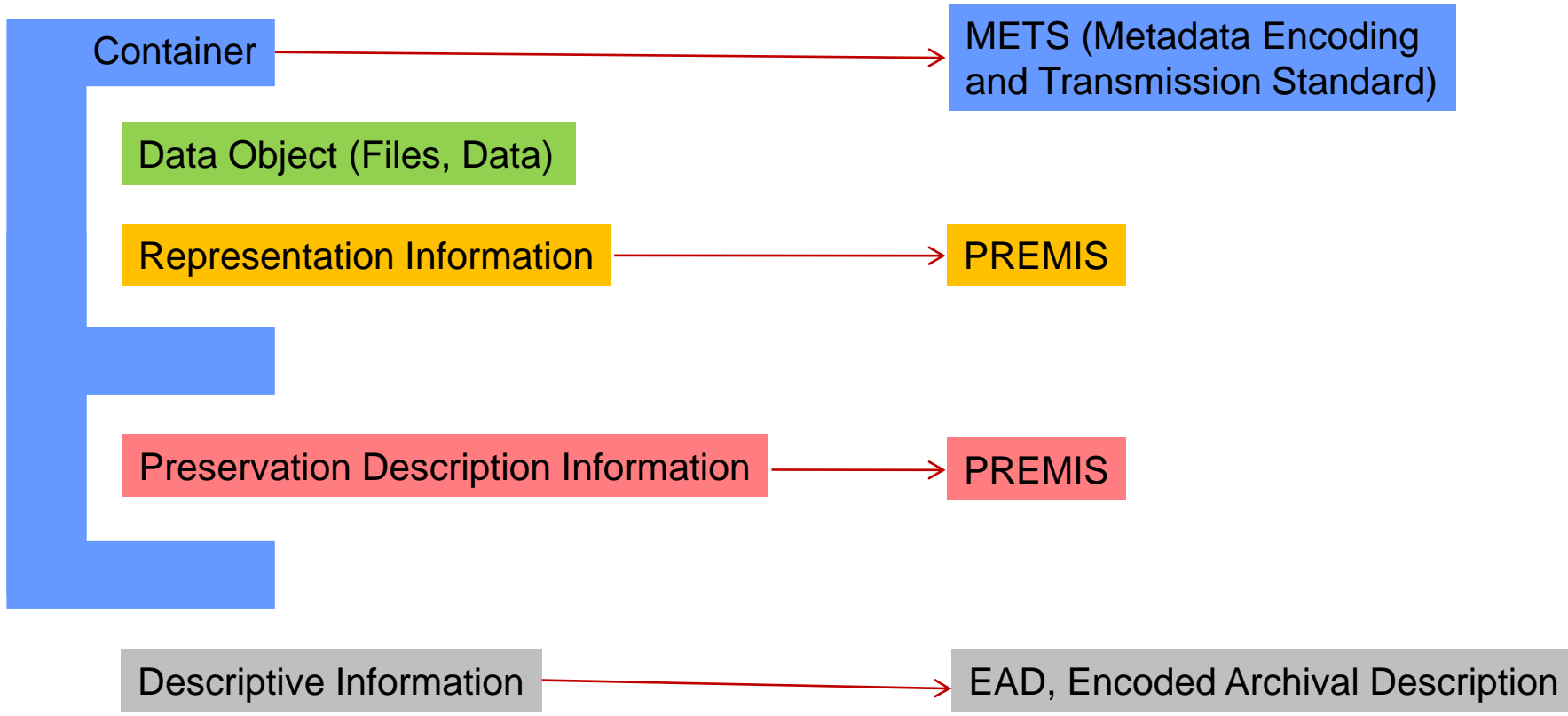


The history: Matterhorn METS

- Matterhorn METS is a metadata model that covers the whole OAIS information model (and not only the descriptive metadata).
- A toolkit consisting of
 - A detailed Metadata-Map
 - The Matterhorn METS Profile registered 2012 at the Library of Congress (LoC)
<https://www.loc.gov/standards/mets/profiles/00000041.xml>
 - Open Source-Software (docuteam packer, docuteam feeder)
<https://docs.docuteam.ch/>
- Use of XML for serialization
- Matterhorn METS is **widely used** nowadays. >70 institutions in Switzerland, France, Germany, Austria rely their digital archiving on it.



Matterhorn METS

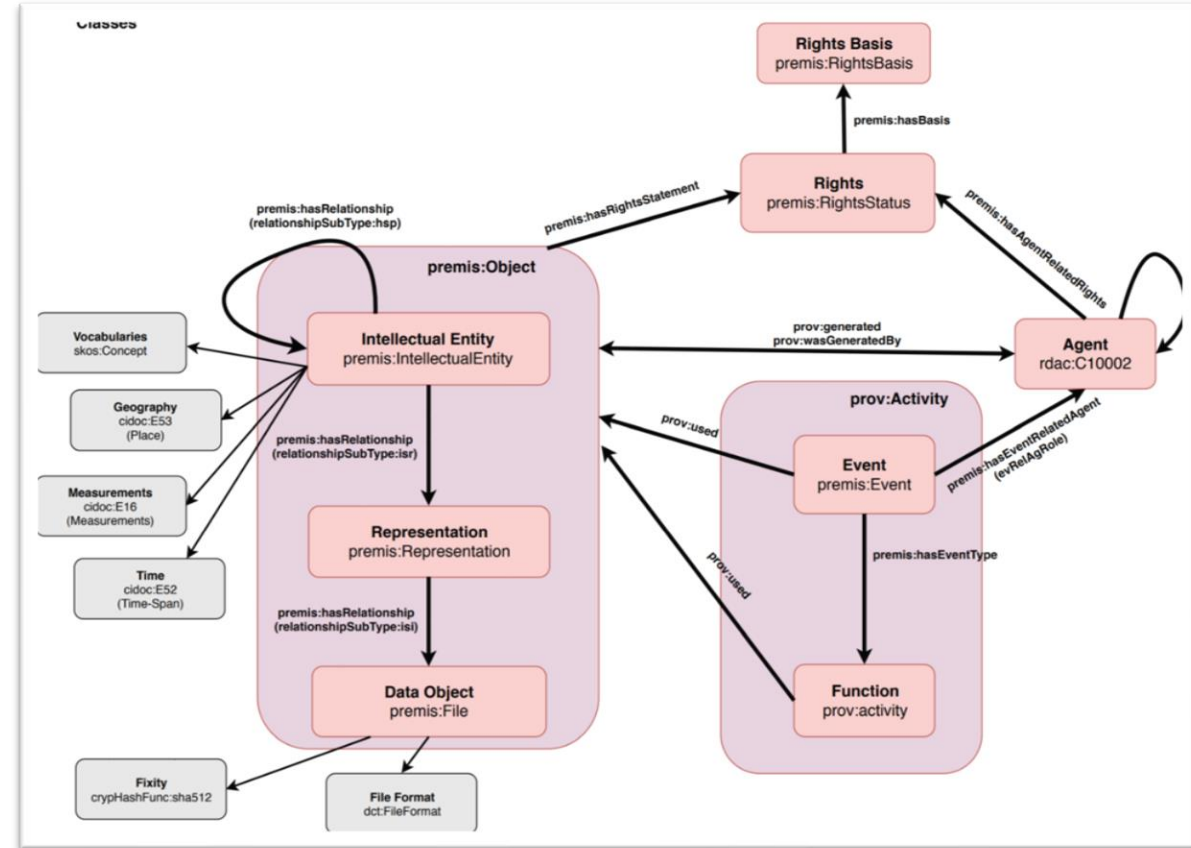


The challenges with Matterhorn METS and the new design decisions

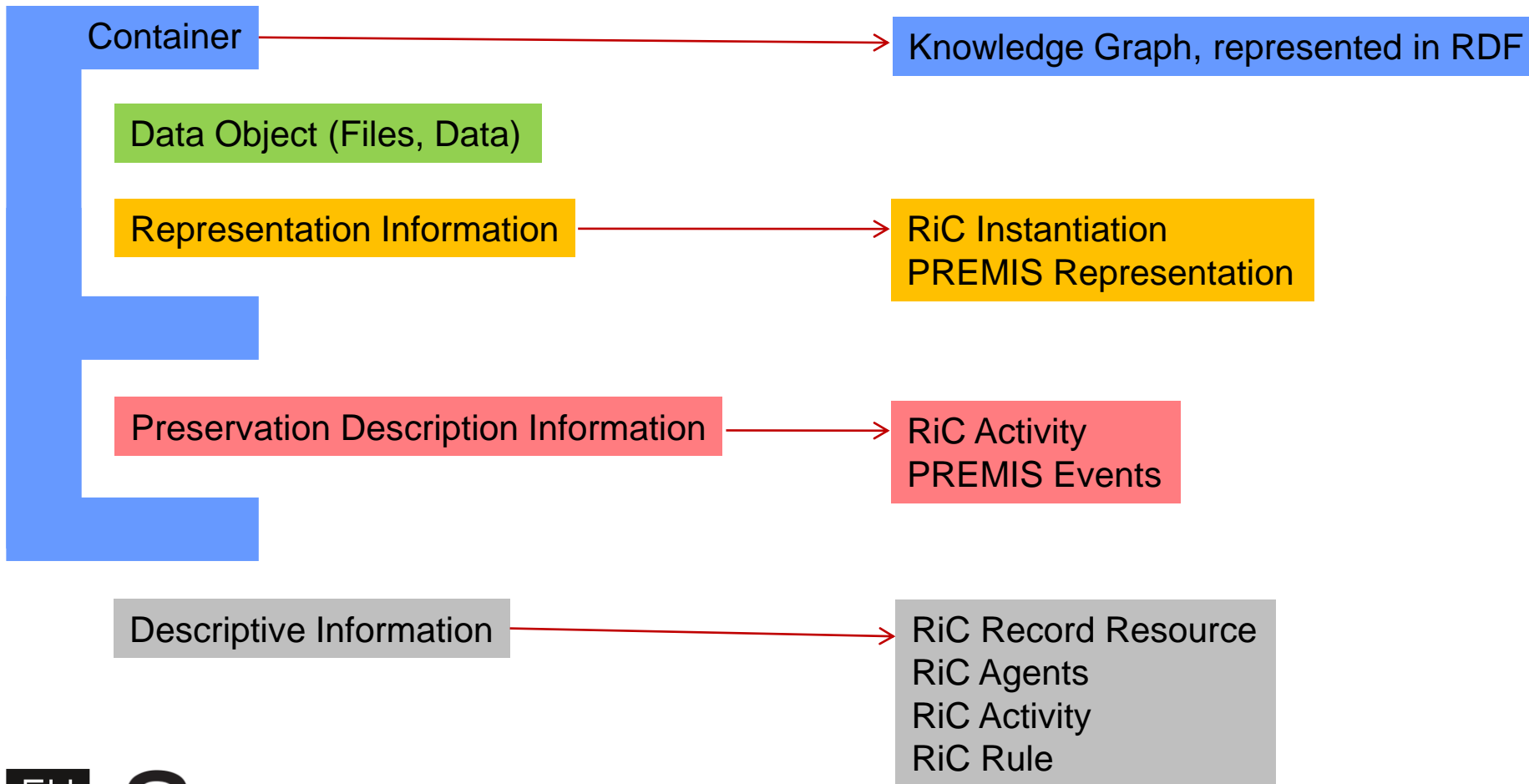
- The existing ICA-standards lack linking between each other. This made a rich contextualization of digital objects difficult
- XML is not sufficient to link these resources
- Semantic technologies are becoming the standard for archival description and library catalogues
- The Matterhorn METS model needed 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)
- **Decision** to use
 - A semantic network instead of a XML-hierarchy to model the digital object
 - Records in Contexts-Ontology (RiC-O) for descriptive metadata
 - PREMIS-Ontology for technical metadata
 - Wherever possible and useful standardized preservation vocabularies from the Library of Congress
<https://id.loc.gov/vocabulary/preservation.html>

From METS to Matterhorn RDF: First Steps

- Our work on Matterhorn RDF started well before a first version of RiC-O was published
- Thus, we chose to work with an **eclectic approach** and combined different existing ontologies to form a new data model: PREMIS, CIDOC, DC, EBUcore, RDA and others
- We stood «on the shoulder of giants», but in the end it was a whole zoo of too many giants!
- For validation, SHACL shapes were very helpful
- As soon as RiC-O v0.1 was released in December 2019, this eclectic approach was no longer necessary and we started to integrate RiC-O.

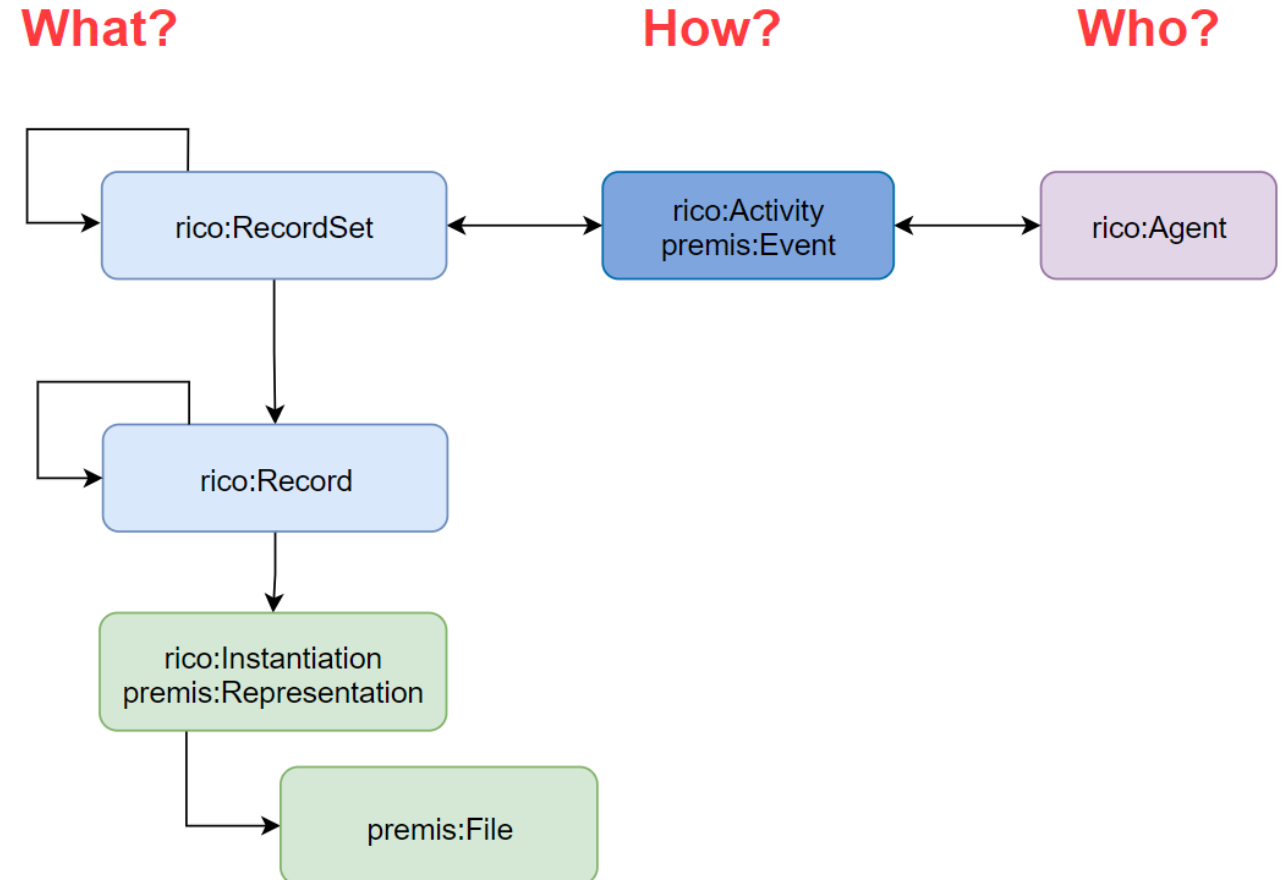


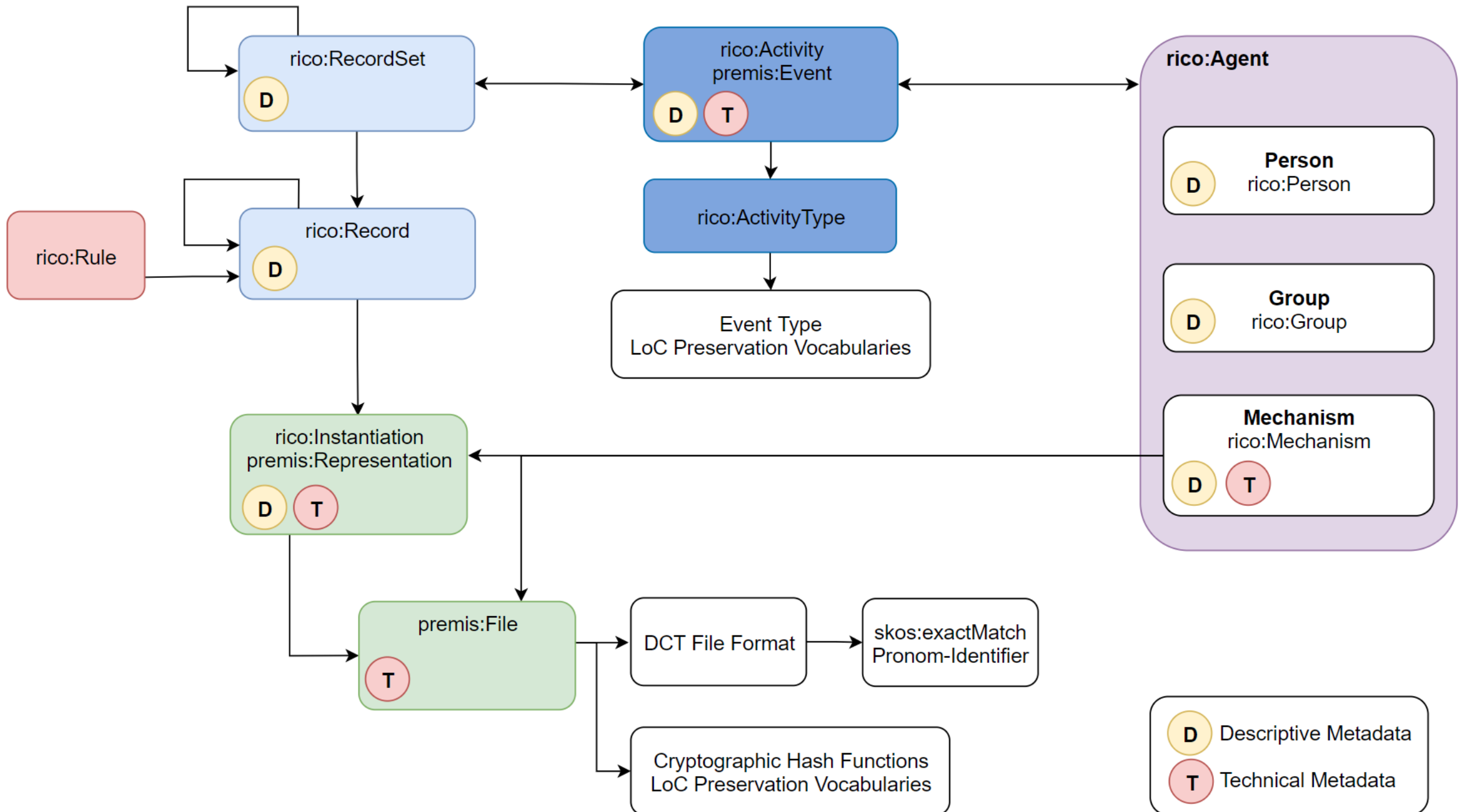
Matterhorn RDF: A Combination of RiC-O and PREMIS



Matterhorn RDF

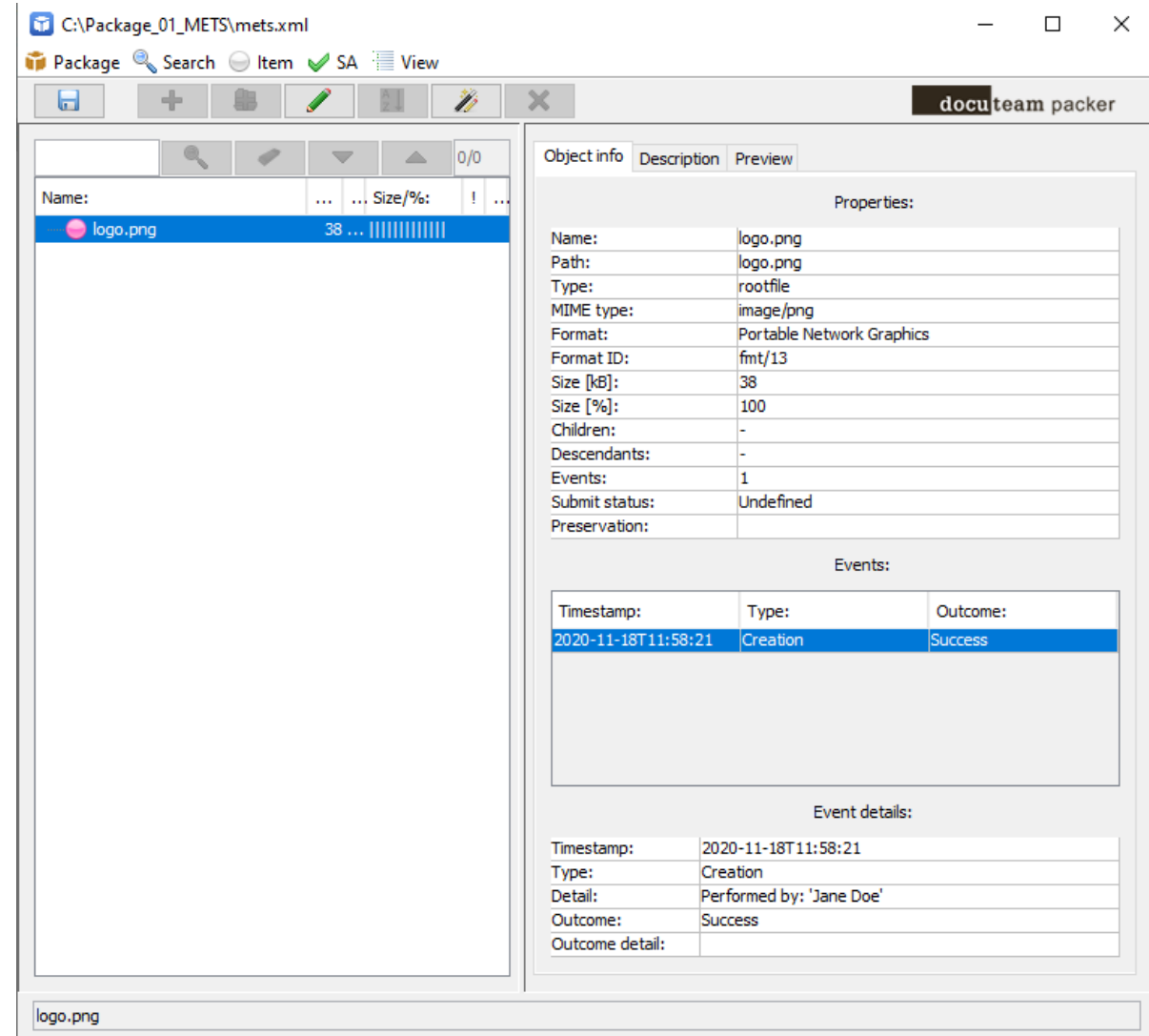
- Matterhorn RDF contains all the necessary metadata to describe and preserve a digital object
- The model provides answers to the questions:
 - **What** does the object contain?
 - **How** has the object been created and preserved?
 - **Who** contributed to the creation and preservation of the object?





Example: From Matterhorn METS to Matterhorn RDF

- Simple digital object containing one PNG-image
- METS-file contains
 - METS header-information
 - Descriptive Metadata Section with EAD
 - Administrative Metadata Section with PREMIS
 - Path to the file
- Download the example:
https://github.com/ICA-EGAD/RiC-O/tree/master/examples/examples_v0-2/Matterhorn-Switzerland



The screenshot shows the 'docu team packer' application window. The title bar indicates the file path is 'C:\Package_01_METS\mets.xml'. The interface includes a toolbar with icons for file operations and a main display area. On the left, a file list shows 'logo.png' with a size of 38 KB. The right pane is divided into 'Object info', 'Description', and 'Preview' tabs. The 'Object info' tab is active, displaying a 'Properties' table and an 'Events' table.

Properties:		
Name:	logo.png	
Path:	logo.png	
Type:	rootfile	
MIME type:	image/png	
Format:	Portable Network Graphics	
Format ID:	fmt/13	
Size [kB]:	38	
Size [%]:	100	
Children:	-	
Descendants:	-	
Events:	1	
Submit status:	Undefined	
Preservation:		

Events:		
Timestamp:	Type:	Outcome:
2020-11-18T11:58:21	Creation	Success

Event details:	
Timestamp:	2020-11-18T11:58:21
Type:	Creation
Detail:	Performed by: 'Jane Doe'
Outcome:	Success
Outcome detail:	

METS (XML)

```
<?xml version="1.0" encoding="UTF-8"?>
<METS:mets xmlns:METS="http://www.loc.gov/METS/"
  xmlns:EAD="urn:isbn:1-931666-22-9"
  xmlns:PREMIS="info:lc/xmlns/premis-v2"
  xmlns:OAI_DC="http://www.openarchives.org/OAI/2.0/oai_dc/"
  xmlns:DC="http://purl.org/dc/elements/1.1/"
  xmlns:xlink="http://www.w3.org/1999/xlink"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.loc.gov/METS/ mets.xsd">
  <METS:metsHdr CREATEDATE="2020-11-18T11:58:20" LASTMODDATE="2020-11-18T11:58:20">
    <METS:agent ROLE="CREATOR" TYPE="INDIVIDUAL">
      <METS:name>t.wildi</METS:name>
    </METS:agent>
  </METS:metsHdr>
  <METS:dmdSec ID="_20201118115821577">...
  </METS:dmdSec>
  <METS:amdSec>...
  </METS:amdSec>
  <METS:fileSec>
    <METS:fileGrp>
      <METS:file ID="_20201118115820184" MIMETYPE="image/png">
        <METS:FLocat LOCTYPE="URL" xlink:href="file:///1" />
      </METS:file>
    </METS:fileGrp>
  </METS:fileSec>
  <METS:structMap>...
  </METS:structMap>
</METS:mets>
```

RDF (TTL)

```
@prefix crypHashFunc: <http://id.loc.gov/vocabulary/preservation/cryptographicHashFunctions/> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix ebucore: <http://www.ebu.ch/metadata/ontologies/ebucore/ebucore#> .
@prefix evttype: <http://id.loc.gov/vocabulary/preservation/eventType> .
@prefix premis: <http://id.loc.gov/vocabulary/preservation/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix relSubType: <http://id.loc.gov/vocabulary/preservation/relationshipSubType> .
@prefix rico: <https://www.ica.org/standards/RiC/ontology#> .
@prefix skos: <http://www.w3.org/2004/02/skos/> .

# ===== Record
> <_20201118115821577> a rico:Record ;...

# ===== Representation
> <_20201118115820159> a rico:Instantiation, premis:Representation ;...

# ===== File
> <_20201118115821547> a premis:File ;...

# ===== Event
> <_20201118115821558> a rico:Activity, premis:Event ;...

# ===== Identifiers
> <501> a rico:Identifier ;...
> <502> a rico:Identifier ;...
> <503> a rico:Identifier ;...
> <504> a rico:Identifier ;...
> <505> a rico:Identifier ;...
> <_20201118115820159> a rico:Identifier ;...
> <_20201118115821558> a rico:Identifier ;...

# ===== Rule
> <700> a rico:Rule ;...
> <701> a rico:Rule ;...
> <702> a rico:Rule ;...

# ===== Agent
> <100> a rico:Mechanism ;...
> <101> a rico:Person ;...

# ===== File Format
> <PNG> a dct:FileFormat ;...

# ===== Event types
> <creation> a evttype:cre ;...
```

Conclusion and Next Steps

- With the help of a semantic network and RDF, the shortcomings of a XML-based model could be eliminated.
- Richer and more complex contextualization (descriptive and technical) is now possible.
- RiC and PREMIS are based on similar principles, but each with a different focus. The two standards complement each other ideally.
- With RiC-O v0.2 and RiC-CM v0.2, RiC is now sufficiently mature to be used in practice.
- Next steps:
 - Our datamodel should be useful not only for archives, but libraries and museums alike. Thus, we plan to do **crosswalks and/or integrations** of other standards and ontologies.
 - The implementation of the data model in **projects and software tools** will deliver experience that is helpful to improve the model where necessary.

Fachhochschule Graubünden
Pulvermühlestrasse 57
7000 Chur
T +41 81 286 24 24
info@fhgr.ch

Thank you very much for your attention.

Fachhochschule Graubünden
Scuola universitaria professionale dei Grigioni
Scola universitaria professionala dal Grischun
University of Applied Sciences of the Grisons

swissuniversities

