Using RiC As Aggregation Format For Cultural Heritage Data In Switzerland

ICA Conference Rome, 22.9.2022
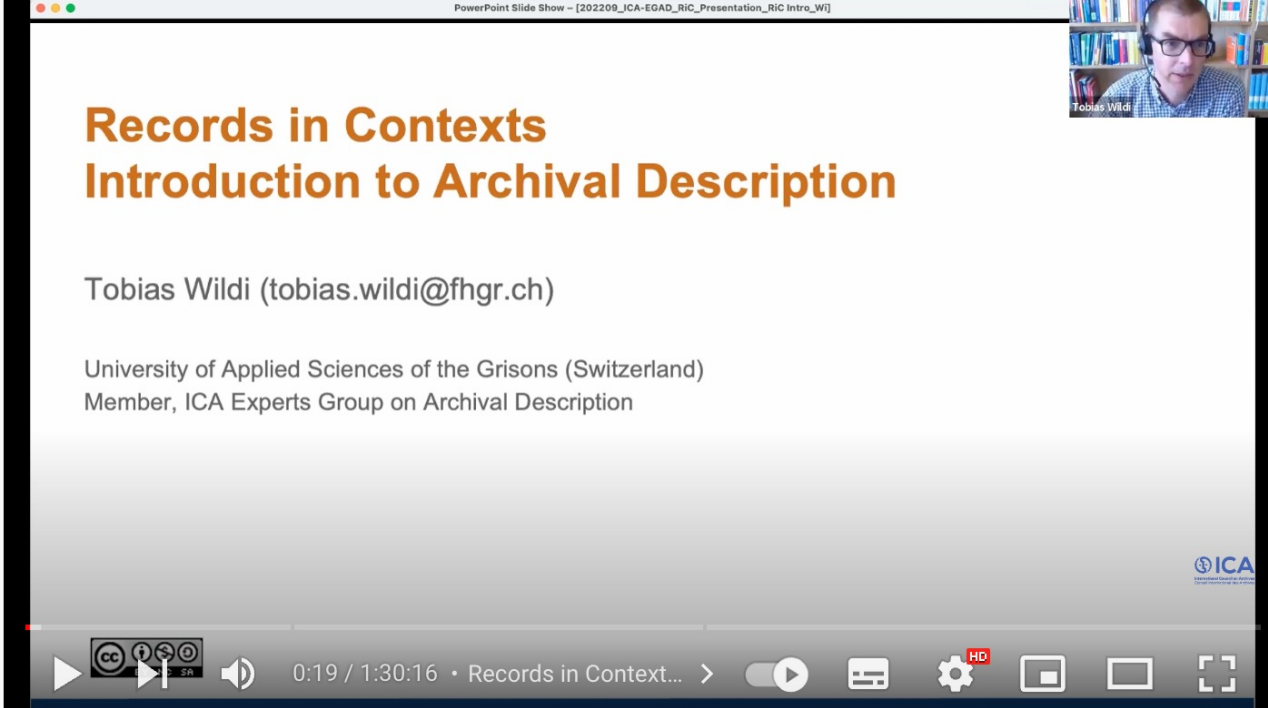
Tobias Wildi, tobias.wildi@fhgr.ch

This talk is not an introduction to RiC…

…but you can find a new introduction to RiC on the ICA Youtube channel

Tobias Wildi	Introduction
Bill Stockting	RiC-CM
Florence Clavaud	RiC-O

https://www.youtube.com/watch?v=oHG_pupre8w

The Swiss Research Data aggregator: Connectome

- An ecosystem of open linked research data (ORD) in Switzerland following the FAIR-principles (FAIR = Findable, Accessible, Identifiable, Reusable)
- Aggregation of metadata from heterogeneous environments
- Run by SWITCH (www.switch.ch) with partners in cultural heritage and research institutions all over Switzerland
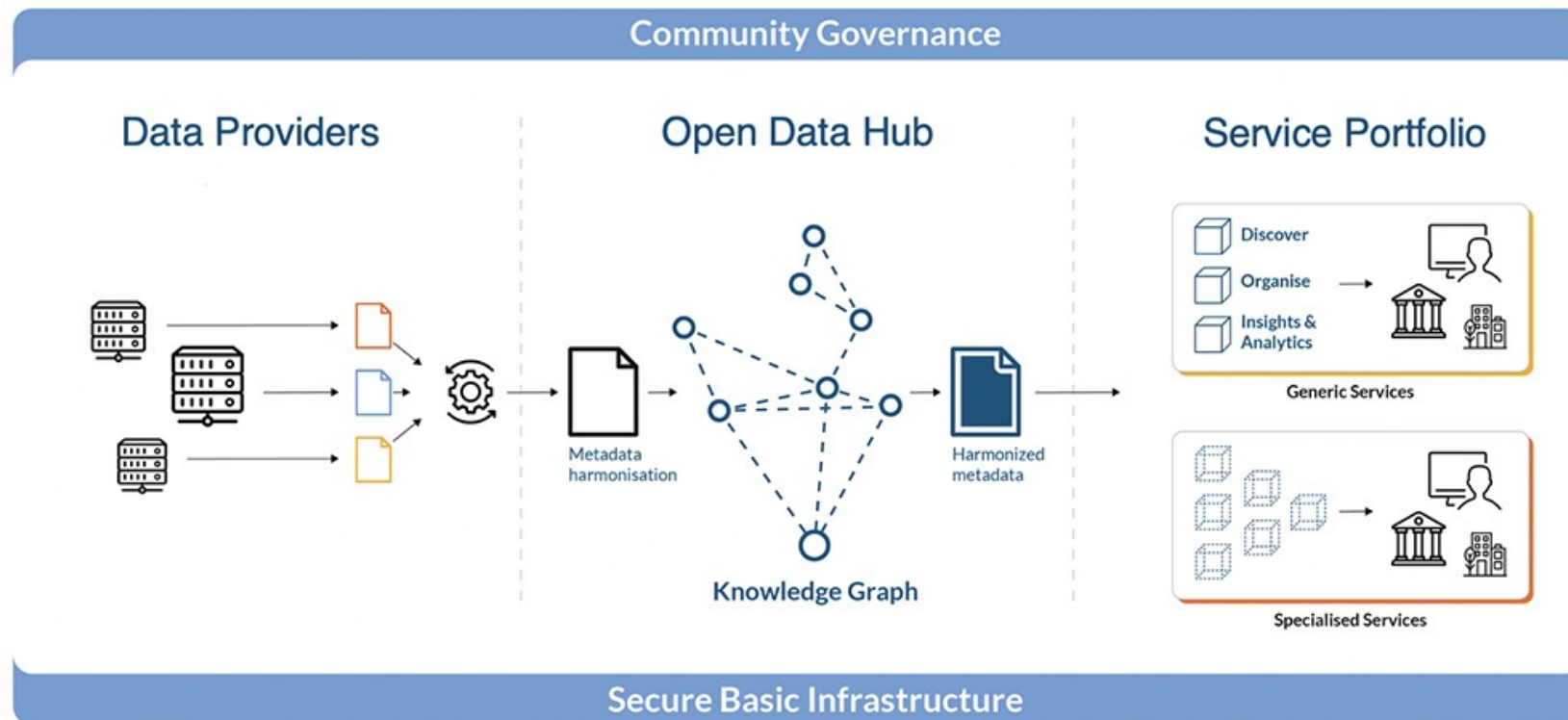
The Use Case: Bringing Patrinum to Connectome

Patrinum from the canton de Vaud (CH)

- Digital Legal Deposit

- Manuscript Collection

- Institutional Repository

- Both library and archival collections

- Internally based on MARC, where as ISAD(G) is encoded in MARC

- Metadata can be harvested via OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting)

https://patrinum.ch/

Records in Contexts as intermediary format for aggregating cultural heritage data

- For cultural heritage data from archives, museums and libraries, RiC is used as intermediate format
- Resources from different repositories are semantically enriched and linked together in the Connectome Knowledge Graph so that service providers can leverage open research data.
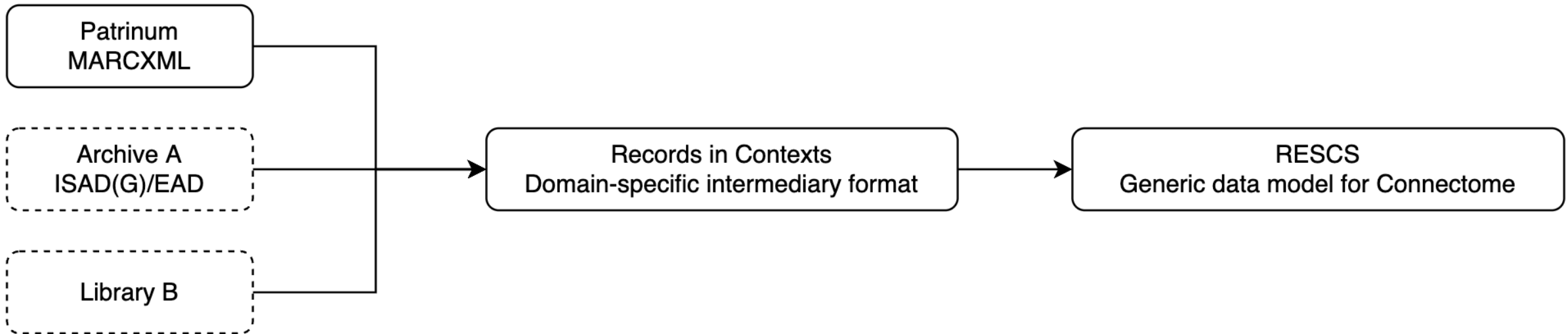
The Mapping Definition

MARC-fields are mapped to RiC-Entities and Attributes (conceptually)

Mapping-definition is written in a declarative language: RML.io

MARC Fields	RiC Entities
00X: Control Fields	rico:Record rico:identifier
041: Language	rico:Language
100, 110: Personal Name, Corporate Name	rico:Person
111: Meeting, Conference	rico:Group
20X-24X: Title fields	rico:Activity
260: Publisher	rico:Title
30X: Physical Description	rico:Group
336: Content Type	rico:Record rico:physicalCharacteristics
340: Physical Medium	rico:PhysicalMedium
347: File Format	rico:CarrierType
351c: Hierarchical Level	rico:Record rico:type

FH
GR

The mapping pipeline based on RDF Mapping language (RML)

Declarative RML mapping definition
https://rml.io/

| Patrinum MARC | → | RML Processor CARML | → | RiC TTL | → | RESCS |

Input: MARCXML is harvested via OAI-PMH

Mapping is performed based on the mapping definition with the help of a mapping-processor

Output: RDF-files in RiC that can be further processed or directly be used

6

Input: MARCXML harvested via OAI-PMH

```xml
 1  <Records xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
 2
 3  <record xmlns="http://www.openarchives.org/OAI/2.0/"><header status=""><identifier>o
 4          <marc:leader>\\\\\nam\a22\\\\\7\\4500</marc:leader>
 5      <marc:controlfield tag="001">303337</marc:controlfield>
 6      <marc:controlfield tag="005">20220312004203.0</marc:controlfield>
 7      <marc:datafield tag="037" ind1=" " ind2=" ">
 8        <marc:subfield code="a">ISADG</marc:subfield>
 9      </marc:datafield>
10      <marc:datafield tag="084" ind1=" " ind2=" ">
11        <marc:subfield code="a">ASM-O-7-1quinquies</marc:subfield>
12      </marc:datafield>
13      <marc:datafield tag="245" ind1="0" ind2=" ">
14        <marc:subfield code="a">Kantonal Chor Aufführungen LA 1939 diverse Anmeldungen f
15        <marc:subfield code="b">1939</marc:subfield>
16      </marc:datafield>
```

Output: RiC in TTL-Format

MARCXML:

```
...
<record>
  <controlfield tag="001">173427</controlfield>
  <controlfield tag="005">20211007005827.0</controlfield>
```
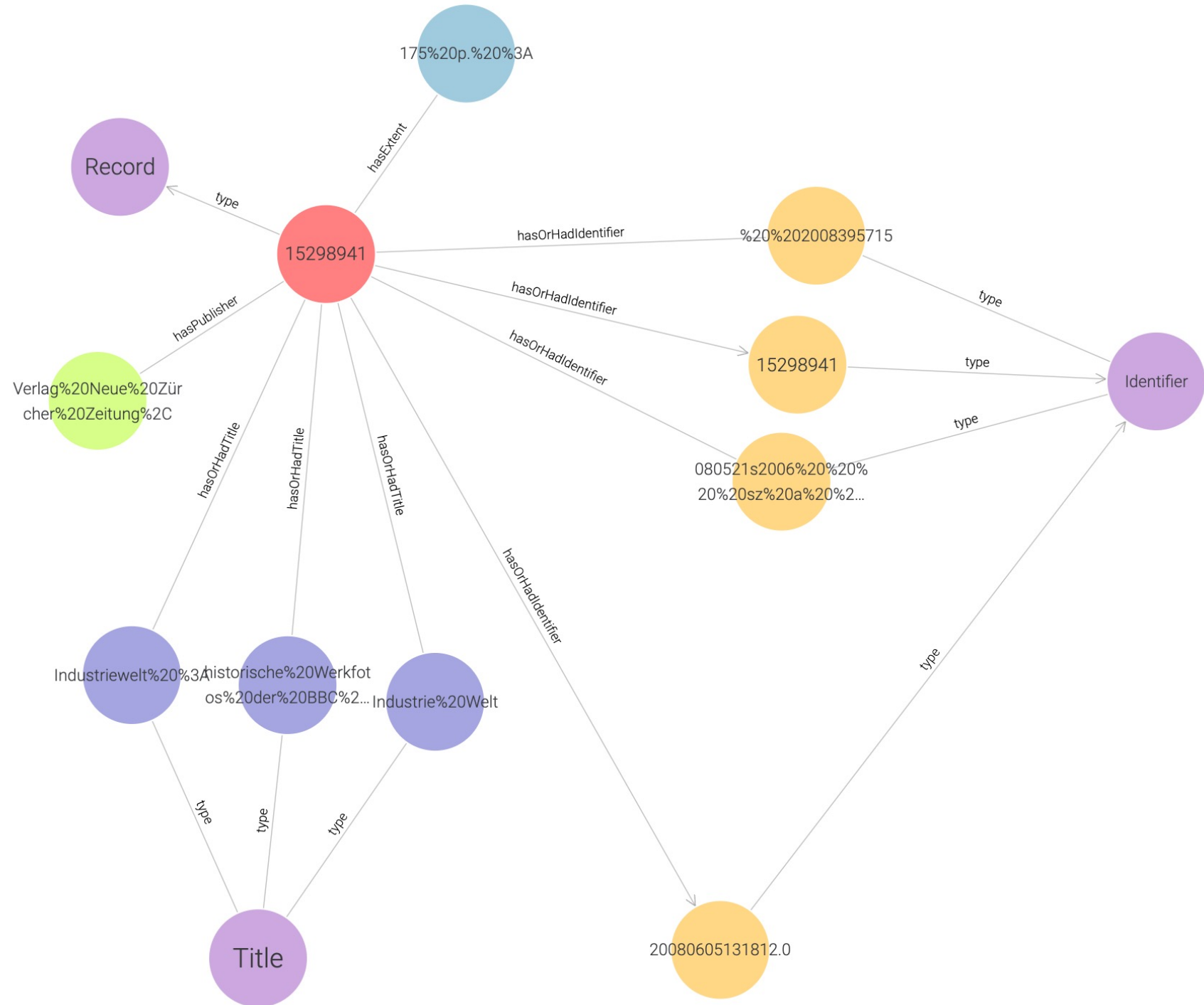
RiC in TTL:

```
<https://data.connectome.ch/Record/173427>
    <https://www.ica.org/standards/RiC/ontology#identifier>
        "173427" .
<https://data.connectome.ch/Record/173427>
    <https://www.ica.org/standards/RiC/ontology#identifier>
        "20211007005827.0" .
```

MARCXML in RiC – visualized

FH GR

Data reuse in Social Science and Humanities

The Research Data Connectome will facilitate and accelerate data reuse in the social sciences and humanities. To find out more about current data research practices in these communities we commissioned the SWITCH Innovation Lab "Repositories & Data Quality". We asked Nicolai Hauf, ZHAW and Martin Jaekel, ZHAW, about their findings and what they could mean for the next steps of building the Research Data Connectome.

Published on 30.04.2021

SWITCH: What was the aim of this SWITCH Innovation Lab?

Nicolai Hauf: We wanted to analyse the reuse of existing research data in the disciplines of Social Sciences and Humanities (SSH). Our main aim was to identify the relevant data sources for researchers in Switzerland. For this reason, we developed a survey investigating the location of existing valuable data, the purpose for reusing this material and the selection criteria researchers use when deciding on data for their own use.

SWITCH: Which data providers are the most frequently named sources?

Nicolai Hauf: Many participants named FORSbase, the digital research information and data access portal for social science studies in Switzerland. The Federal Statistical Office (FSO) is another very important source for data on many different research topics. Other important mentions are the European Social Survey (ESS) and the GESIS data archive. Interesting is also the wide variety of small data providers that were mentioned.

Dr Martin Jaekel

Dr Martin Jaekel is the Head of R&D Unit at the Zurich University of Applied Sciences and has been co-leading the swissuniversities working group developing the Swiss ORD-Strategy and Action Plan (the *analysis report* ⧉, providing a

https://www.switch.ch/stories/Data-reuse-in-Social-Science-and-Humanities/

10

Findings

- The goal of the Connectome aggregator is to build a sustainable platform to make cultural heritage data, among others, reusable for the humanities and social sciences.

- In this context, RiC is a powerful format that can aggregate metadata from heterogeneous cultural heritage environments

- Source metadata formats often evolve over time. Thus, metadata mappings must be flexible

- Implementing the mappings with declarative, configurable technologies like RML makes refactoring of mapping much easier than having to change hard coded scripts and sofware.

FH
GR

Thank you for your attention.