Linked Cultural Heritage Data? FAIR Enough!

58th CIDOC CRM & 51st FRBR/LRMoo CRM

Inês Koch
Carla Teixeira Lopes
Cristina Ribeiro
María Poveda Villalón
Mariano Rico

This work is financed by National Funds through the Portuguese funding agency, FCT – Fundação para a Ciência e a Tecnologia within the research grant 2020.08755.BD.
Inês Koch

Ph.D. candidate - Doctoral Programme in Digital Media
EPISA Project - Entity and Property Inference for Semantic Archives

Master in Information Science

Porto, Portugal
Outline

- Background
- EPISA
- ArchOnto
- Archival Records Representation
- EPISA Platform
Background
Linked Data Models

- **CIDOC CRM**
  - Created in the scope of museums by the CIDOC
  - Developed by the CRM SIG (CRM Special Interest Group)

- **RiC-CM**
  - Created in the scope of archives by the ICA
  - Developed by the EGAD (Expert Group on Archival Description)

- **FRBRoo**
  - Created in the scope of libraries by the IFLA
Linked Data Models

- Europeana Data Model (EDM)
  - Created in the scope of Europeana Project

- DBpedia
  - A crowd-sourced community effort to extract structured content from the information created in various Wikimedia projects

- Wikidata
  - Central storage repository that can be accessed by others, such as the wikis maintained by the Wikimedia Foundation
FAIR Principles

- The FAIR Principles are sources for the requirements that data and metadata need to meet and have the necessary modular structure.
  - Findable
  - Accessible
  - Interoperable
  - Reusable
EPISA

Entity and Property Inference for Semantic Archives
EPISA Project

- Part of the ongoing renewal of the The Portuguese National Archives’ existing data infrastructure

- Goals:
  - Develop a prototype for an open-source knowledge graph platform representing archival information on a linked data model.
  - Find ways to guarantee the migration of contents stored according to ICA (International Council on Archives) standards to an ontology-based model, the CIDOC CRM (Conceptual Reference Model).
From DigitArq to DigitArq+
Proposed evolution of archival description
Atomization of an ISAD(G) record in linked data
ArchOnto

An extension of CIDOC CRM for archives
ArchOnto

- Is composed by 5 different ontologies:
  - CIDOC CRM
  - DataObject
  - N-ary
  - ISAD Ontology
  - Link2DataObject

- Use the prefixes:
  - ARE – Archival Entity, ex.: ARE1 Level of Description.
  - ARP – Archival Property, ex.: APR12 has level of description.
## ArchOnto - Controlled vocabularies

<table>
<thead>
<tr>
<th>Class</th>
<th>Vocabulary</th>
<th>Example Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARE1 Level of Description</td>
<td>Level of description</td>
<td>Fonds; Series; Section; File; Item</td>
</tr>
<tr>
<td>ARE2 Formal Title</td>
<td>Title Type</td>
<td>Formal, Supplied</td>
</tr>
<tr>
<td>ARE3 Supplied Title</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARE5 Identifier Type</td>
<td>Identifier of collective person/group</td>
<td>PT; VCT; AGH01; 161016; ADLSB; 60084892; FT-LIBM</td>
</tr>
<tr>
<td>ARE6 Date Type</td>
<td>Type of time period</td>
<td>Exact dates; Inferred dates; Predominant dates</td>
</tr>
<tr>
<td>ARE9 Date Certainty</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARE7 Name Type</td>
<td>Type of name of collective person/group</td>
<td>Authorized form of name; Another form of the name; Parallel name form</td>
</tr>
<tr>
<td>ARE8 Role Type</td>
<td>Role played</td>
<td>Producer; Material Author; Recipient</td>
</tr>
<tr>
<td>ARE11 Documentary Typology</td>
<td>Documentary Typology</td>
<td>Certificate; Income book; Patent</td>
</tr>
<tr>
<td>ARE13 Subject Type</td>
<td>Subject</td>
<td>Education; Science; Law; Management</td>
</tr>
<tr>
<td>ARE14 Place Type</td>
<td>Type of jurisdictional entity</td>
<td>Ocean; Archipelago; Mountain range; Country; District</td>
</tr>
<tr>
<td>ARE15 Acquisition Type</td>
<td>Transfer of Custody / Acquisition Identifier</td>
<td>Purchase; Giving; Donation; Deposit; Swap; Legacy; Reintegration; Transfer</td>
</tr>
<tr>
<td>ARE16 Event Type</td>
<td>Event Type</td>
<td>Evaluation; Expertise; Financial management</td>
</tr>
<tr>
<td>E56 Language</td>
<td>Language Identifier</td>
<td>Portuguese; Latin; French; Greek</td>
</tr>
<tr>
<td>E57 Material</td>
<td>Support</td>
<td>Paper; Parchment; Photosensitive film</td>
</tr>
<tr>
<td>E58 Measurement Unit</td>
<td>Measurement Unit</td>
<td>Centimeter; Gram; Byte; Minute; Pack</td>
</tr>
<tr>
<td>E59 Currency</td>
<td>Currency</td>
<td>Euro; Dollar; Kwanza</td>
</tr>
</tbody>
</table>
ArchOnto
ArchOnto - CIDOC CRM
ArchOnto - DataObject
ArchOnto - N-ary
ArchOnto - ISAD Ontology
ArchOnto - Link2DataObject
Categorization of Archival Concepts - ArchOnto
Archival Records Representation
DigitArq Record
ArchOnto Representation
Wikidata Representation
ArchOnto + DBpedia + Wikidata Representation
ArchOnto VS DBpedia VS Wikidata
ArchOnto VS DBpedia VS Wikidata
ArchOnto VS DBpedia VS Wikidata

Death Date

E21 Person
Person001

P100 died in E69 Death
Death001

P4 has time-span E52 Time-Span
Time-Span001

P1 is identified by E41 Appellation
Appellation006

L200 has value DOE10 Instant
Instant002

xsd.date
dbo.deathDate

Q215627 Person
Person001

P570 date of death Point in time

Death Place

E21 Person
Person001

P100 died in E67 Birth
Birth001

P7 took place E53 Place
Place001

P1 is identified by E41 Appellation
Appellation005

L200 has value DOE8 String
String002

xsd.string
dbo.deathPlace
dbo.place

Q215627 Person
Person001

P20 place of death Item

xsd.dateTime
ArchOnto VS DBpedia VS Wikidata
ArchOnto VS DBpedia VS Wikidata
EPISA Platform

An archival management system based on linked data technologies
EPISA Platform

The EPISA Platform is the software infrastructure developed to support archival records management based on linked data technologies.

Supports storage, creation, search, access, and navigation over archival records and related entities.

Is based on open-source native linked data technologies.

Organized in two main components:

- **EPISA Server** is responsible for storing, reasoning, managing access, and providing an effective search mechanism over the archival data.

- **EPISA ArchClient** is a web application providing a graphical user interface for archivists to access, manage and describe collections of archival records.
EPISA Platform Docker Environment
User Interface Abstractions

Higher-level abstractions need to be defined at the user interface level.

The abstractions defined are:

- **Records**, representing the collection of archival documents (e.g., fonds, collections, records).
- **Entities**, representing concepts (i.e., persons, places, organizations) mentioned in the records.
- **Events**, allowing the creation of complex structures linking records and entities — e.g., birth and death events, marriages, places of domicile.
Central Information Concepts at the UI Level
EPISA ArchClient

ArchClient

Search for archival records, places, events and more
Linked Cultural Heritage Data? FAIR Enough!

58th CIDOC CRM & 51st FRBR/LRMoo CRM

Inês Koch
Carla Teixeira Lopes
Cristina Ribeiro
María Poveda Villalón
Mariano Rico

This work is financed by National Funds through the Portuguese funding agency, FCT – Fundação para a Ciência e a Tecnologia within the research grant 2020.08755.BD.