EPISA - ENTITY AND PROPERTY INFERENCE FOR SEMANTIC ARCHIVES

55th CIDOC CRM and 48th FRBR CRM

Inês Koch
Carla Teixeira Lopes
Cristina Ribeiro
Sérgio Nunes

This work is financed by National Funds through the Portuguese funding agency, FCT – Fundação para a Ciência e a Tecnologia within project DSAIPA/DS/0023/2018. Inês Koch is also financed by FCT 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

- EPISA Project
- ArchOnto
- EPISA Platform
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).
EPISA team

**INESC TEC:** Cristina Ribeiro, Carla Teixeira Lopes, Gabriel David, Sérgio Nunes, Bruno Giesteira, Inês Koch, Mariana Dias, Cláudia Martins, Tiago Silva, Tiago Gomes, Catarina Pires, Camilla Oliveira, Pedro da Silva

**U.Évora:** Irene Rodrigues, Dora Melo, Davide Varagnolo

**DGLAB:** Ana Maria Rodrigues, Lucília Runa, Maria José Almeida, Cristiane Casaca, Silvestre Lacerda
EPISA tasks

- T1: Project management and dissemination of results
- T2: Linked data model for cultural content
- T3: Semantic migration with relations inference
- T4: Knowledge propagation
- T5: Document mining for automatic metadata records
- T6: Knowledge graph management platform KGMP
- T7: Exploration and querying interface
- T8: User evaluation
From Digitarq to Digitarq+
From Digitarq to Digitarq+
From Digitarq to Digitarq+
From Digitarq to Digitarq+
From Digitarq to Digitarq+
From Digitarq to Digitarq+
From Digitarq to Digitarq+
From Digitarq to Digitarq+
From DigitArq to DigitArq+
Proposed evolution of archival description
Atomization of an ISAD(G) record in linked data

https://digitarq.adevr.arquivos.pt/details?id=1174365
Atomization of an ISAD(G) record in linked data

https://digitarq.adevr.arquivos.pt/details?id=1174365
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
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; 600084982, FT-LiBN</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>E98 Currency</td>
<td>Currency</td>
<td>Euro; Dollar; Kwanza</td>
</tr>
</tbody>
</table>
ArchOnto - CIDOC CRM
ArchOnto - DataObject
ArchOnto - N-ary
ArchOnto - Link2DataObject
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 Homepage
Hierarchical Navigation

- Hierarchical Navigation
- "Um aquário de 2,40x1,00 m"³
- Inventario pelo óbito de Manuel Joaquim Ferreira Lino
- Mosteiro de São Bernardino de Portalegre
- Registo de passaportes deferidos
- Julho da Índia e Mina
  - Auto de petição e justificação para protesto de avaria grossa e rectificação do mesmo em que é autor Acácio José Correia,...
  - Acção civil de fretes em que é autor António Henrique Leal e réu António Esteves Costa
  - Extrato do jornal a bordo do navio sueco "Jøanes de Otemburgo"
  - Procuração de Vasco Saltor Mascarenhas e Mendonça dando poderes a João Manuel da Costa
  - Cartório do escrivão Lino José de Almada Lobo da Torre do Vale
- Julho das Justificações Ultramarinas
  - Auto de sentença de José Borges Leal, Dionísio José, Jacinto Borges Leal, Mariana de Jesus, Joaquim Machado, filhos de...
  - This document does not have descendants
  - Maço 5
  - Oriente
- Desembargado do Paço
- Mosteiro de São Bernardo da Rotária
Search by Keywords (archival records results)
Search by Keywords (entities results)
Single Record Description
Single Record References

Autos de sentença de José Borges Leal, Dionísio José, Jacinto Borges Leal, Mariana de Jesus, Joaquim Machado, filhos de Manuel Leal e de Rita Mariana, naturais e moradores no distrito de Vila da Praia, Ilha Terceira ( Açores)

http://www.episa.inesctec.pt/archonto#document17

Add new reference

Type

Existing references

<table>
<thead>
<tr>
<th>Relationship</th>
<th>Target Node</th>
</tr>
</thead>
<tbody>
<tr>
<td>refers to</td>
<td>The Terceira</td>
</tr>
<tr>
<td>documents</td>
<td>Rita Machado</td>
</tr>
<tr>
<td>documents</td>
<td>Manuel Leal</td>
</tr>
<tr>
<td>documents</td>
<td>José Borges Leal</td>
</tr>
<tr>
<td>documents</td>
<td>Dionísio José</td>
</tr>
<tr>
<td>documents</td>
<td>Jacinta Borges Leal</td>
</tr>
<tr>
<td>documents</td>
<td>Mariana de Jesus</td>
</tr>
</tbody>
</table>
View Record ArchOnto Source
View Entity of Person Type
Development Challenges

- Limited maturity of linked data technologies, i.e. no standard off-the-shelf solutions.
- Lack of tools and resources focused on linked data authoring platforms.
- Lack of examples or guidelines for user interaction in authoring linked data.
- Complex underlying linked data model.
ENTITY AND PROPERTY INFERENCE FOR SEMANTIC ARCHIVES

55th CIDOC CRM and 48th FRBR CRM

Inês Koch
Carla Teixeira Lopes
Cristina Ribeiro
Sérgio Nunes

Follow us:
http://episa.inesctec.pt
https://purl.org/episa/archonto