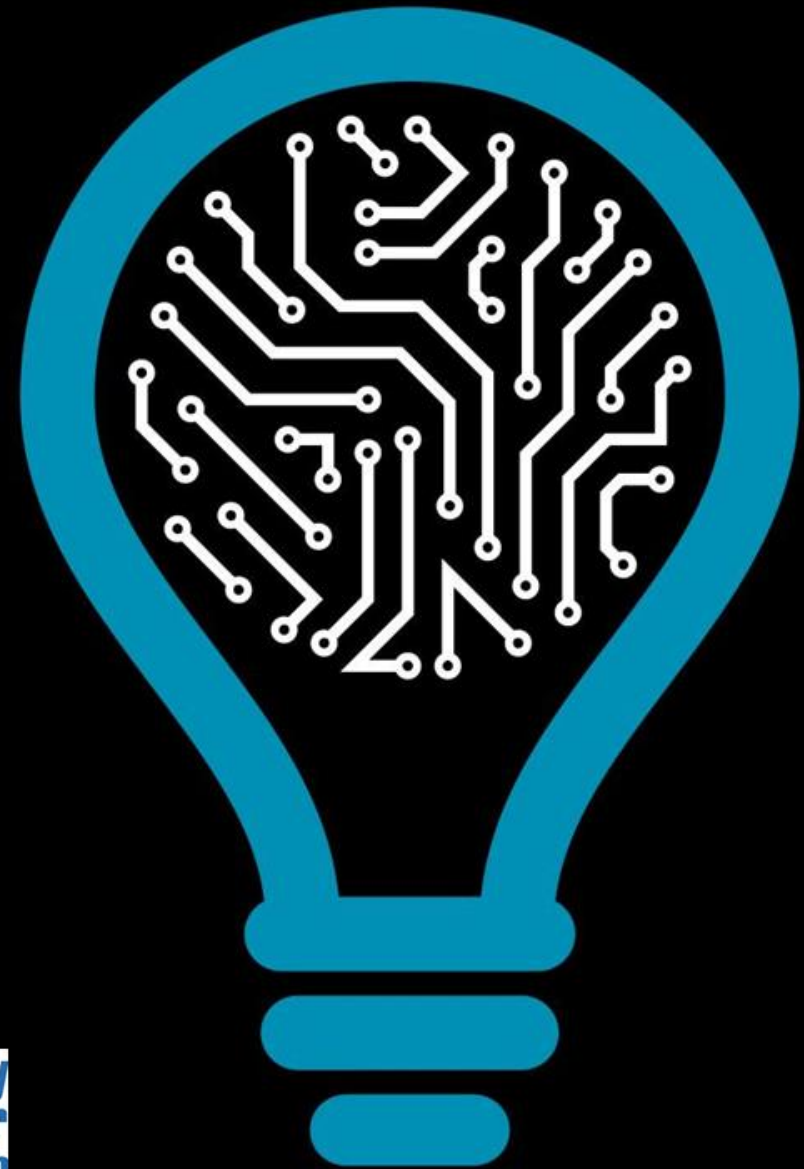# Linked Cultural Heritage Data?
# FAIR Enough!

**58th CIDOC CRM & 51st FRBR/LRMoo CRM**

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

INESCTEC

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Ontology
Engineer
ingGroup

# 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.

  - **F**indable

  - **A**ccessible

  - **I**nteroperable

  - **R**eusable

# 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).
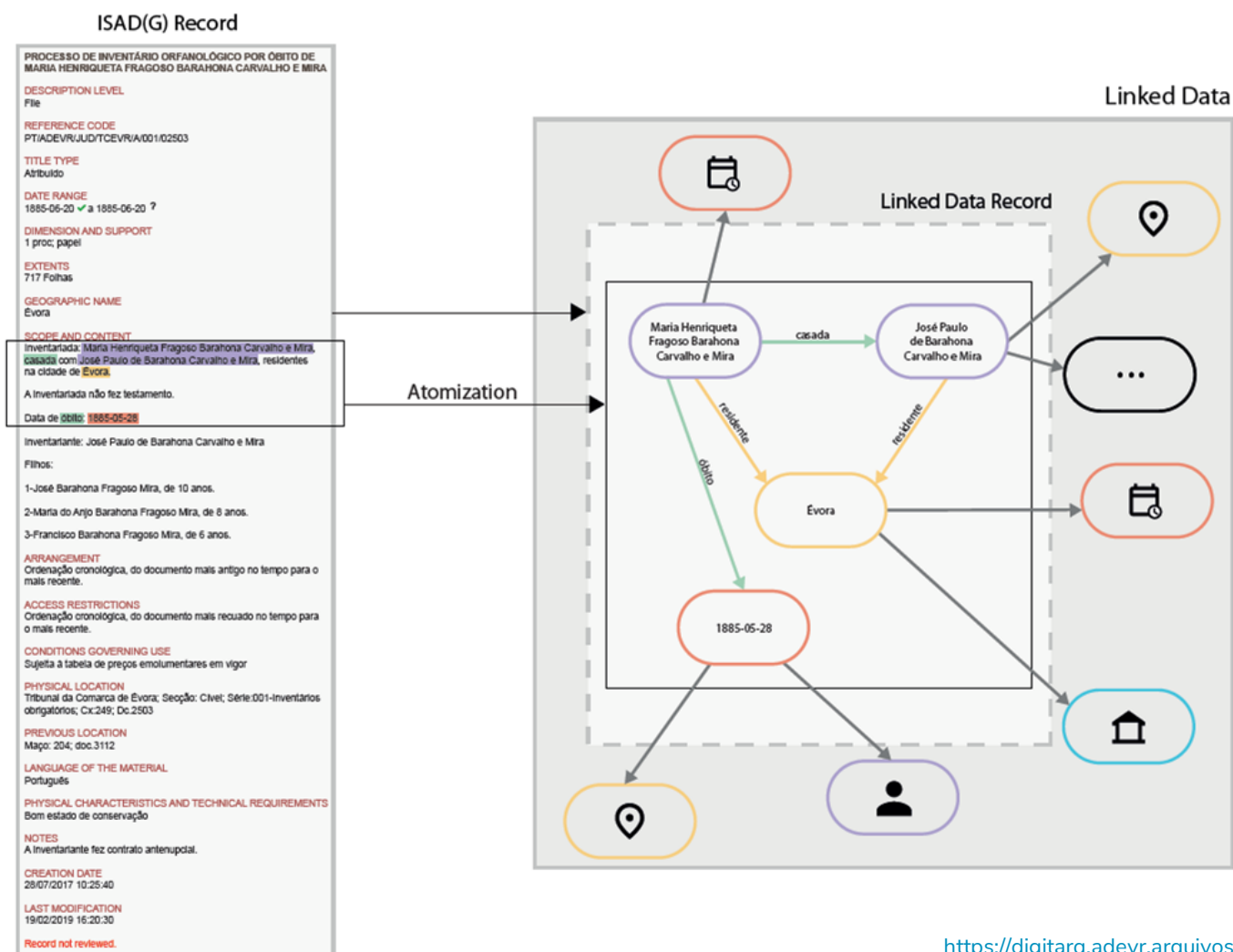
# From Digitarq to Digitarq+
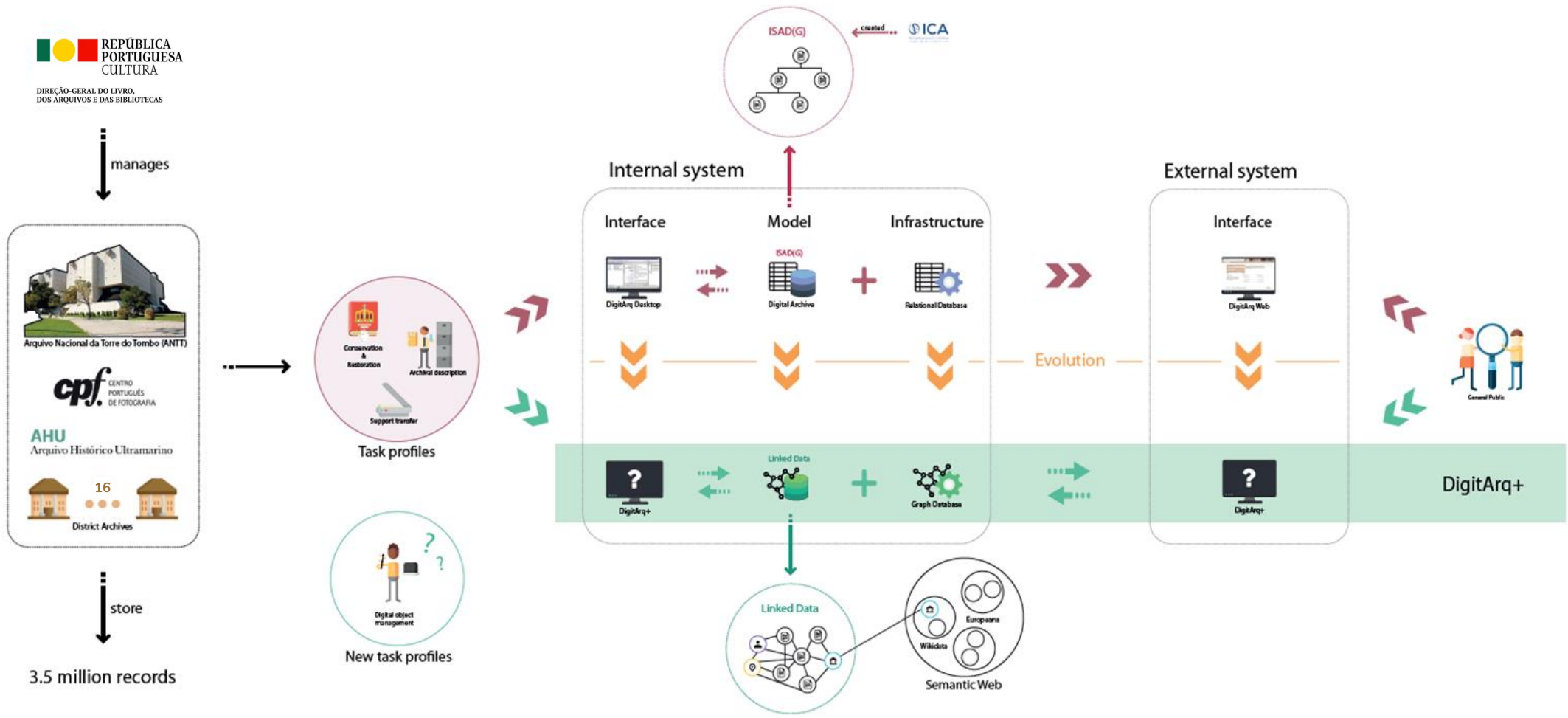
# Proposed evolution of archival description
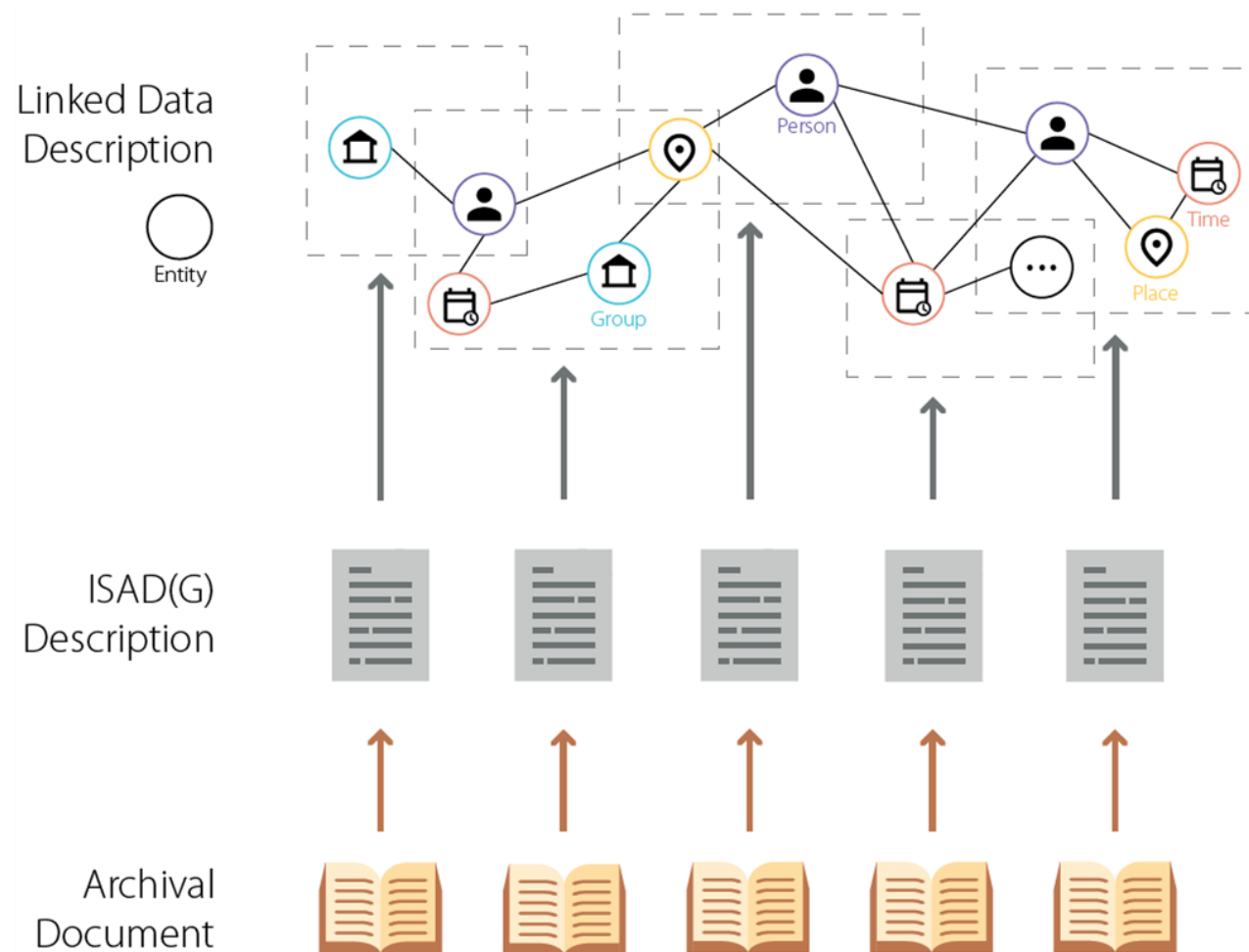
# Atomization of an ISAD(G) record in linked data

12

# ArchOnto

An extension of CIDOC CRM for archives
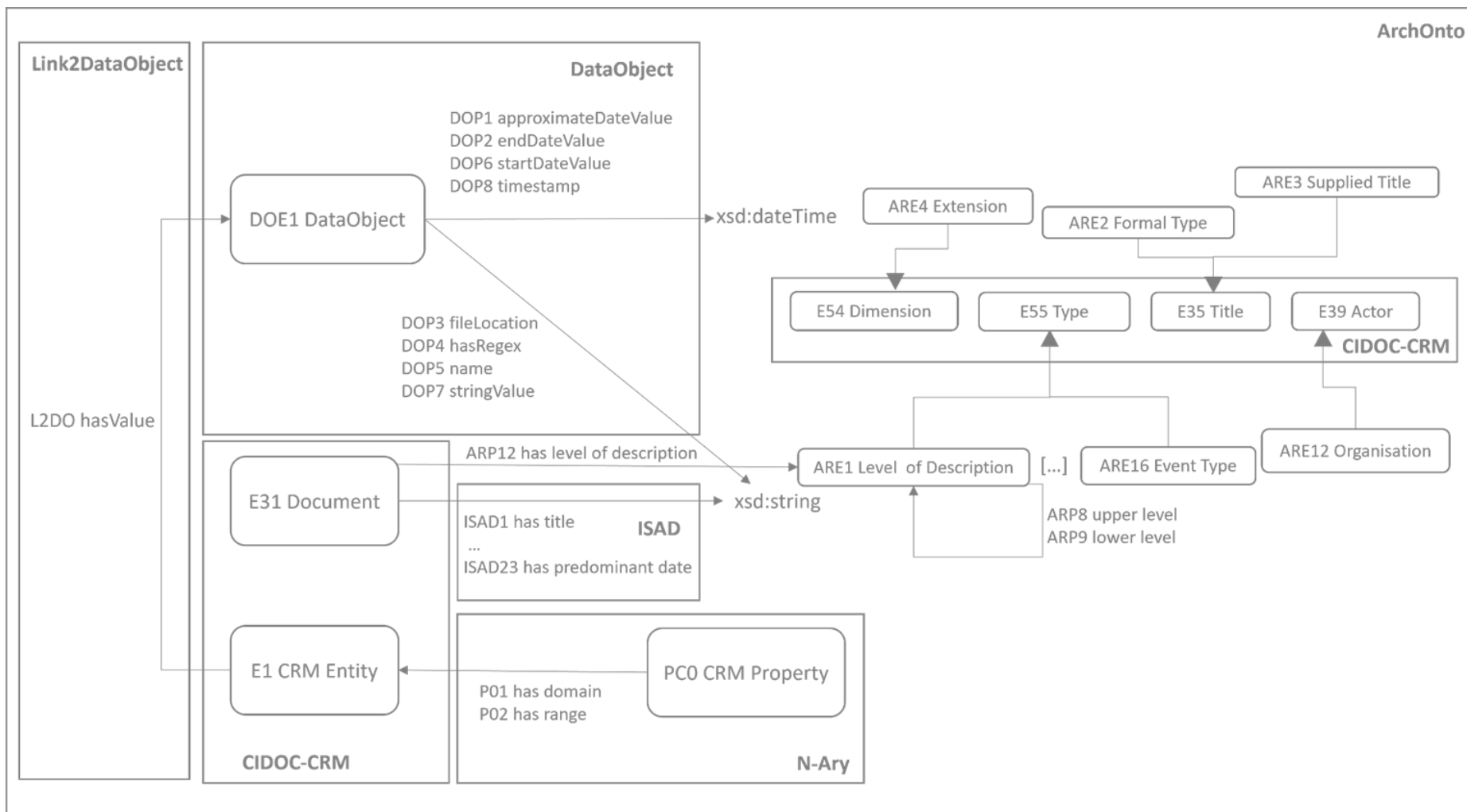
# ArchOnto

- 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

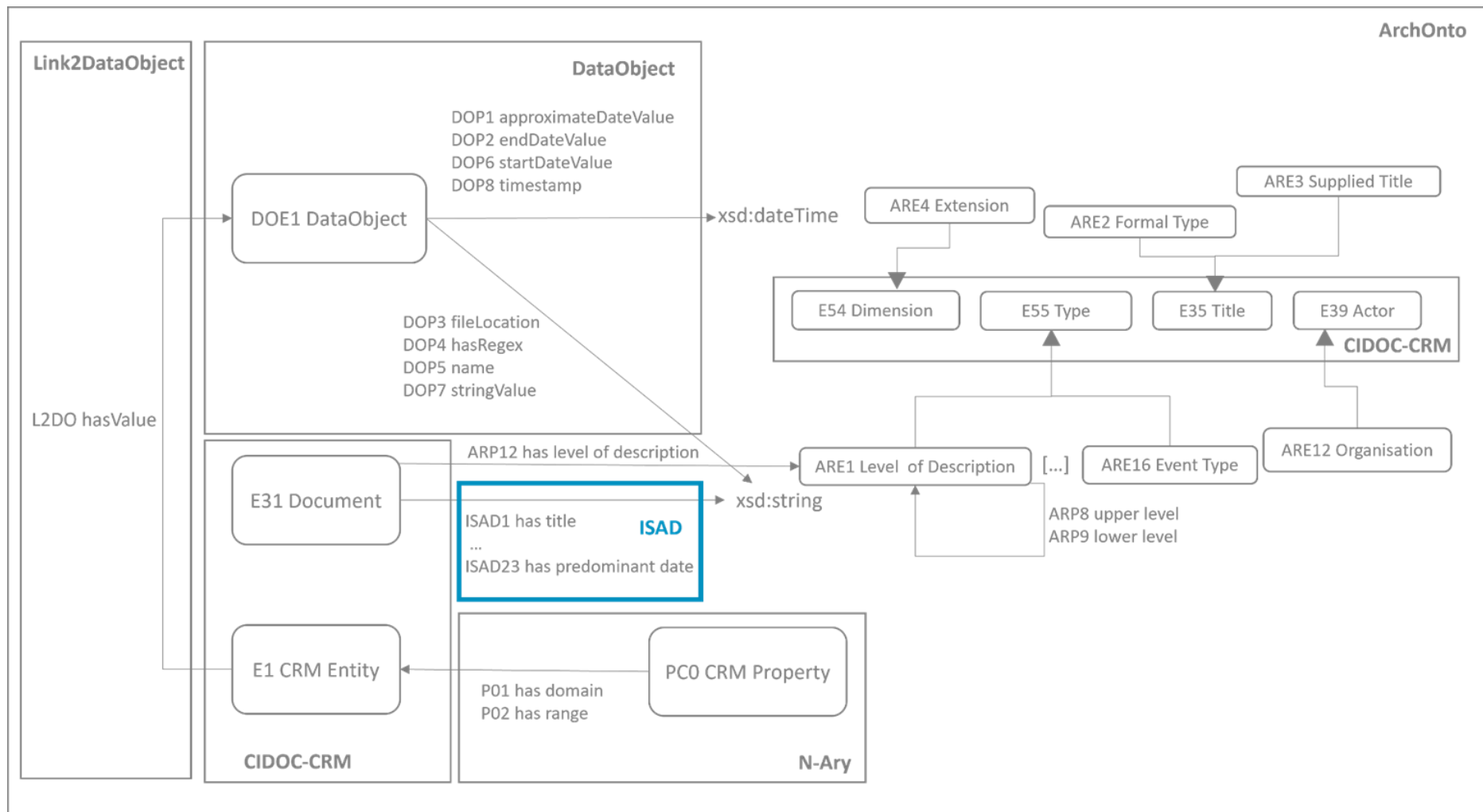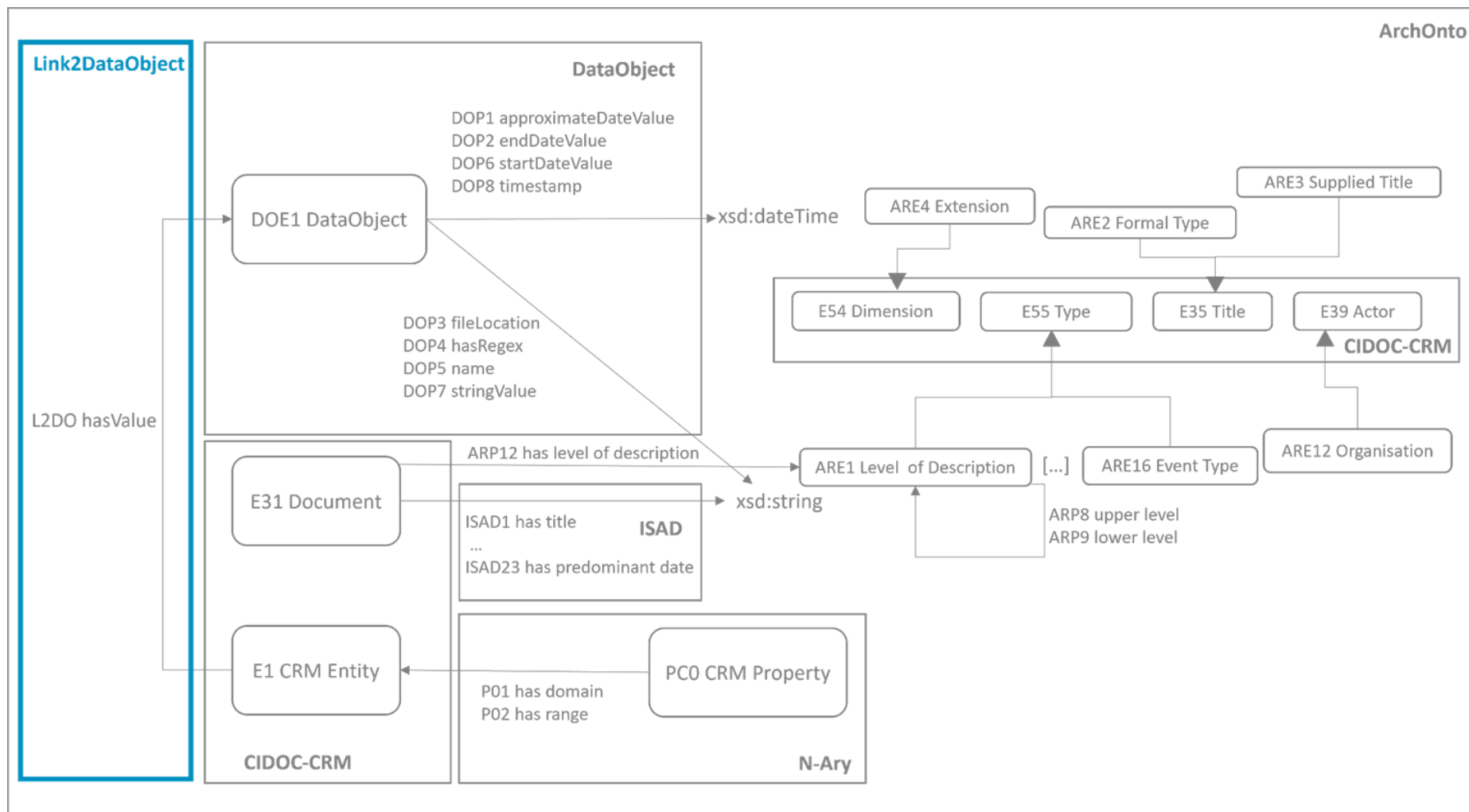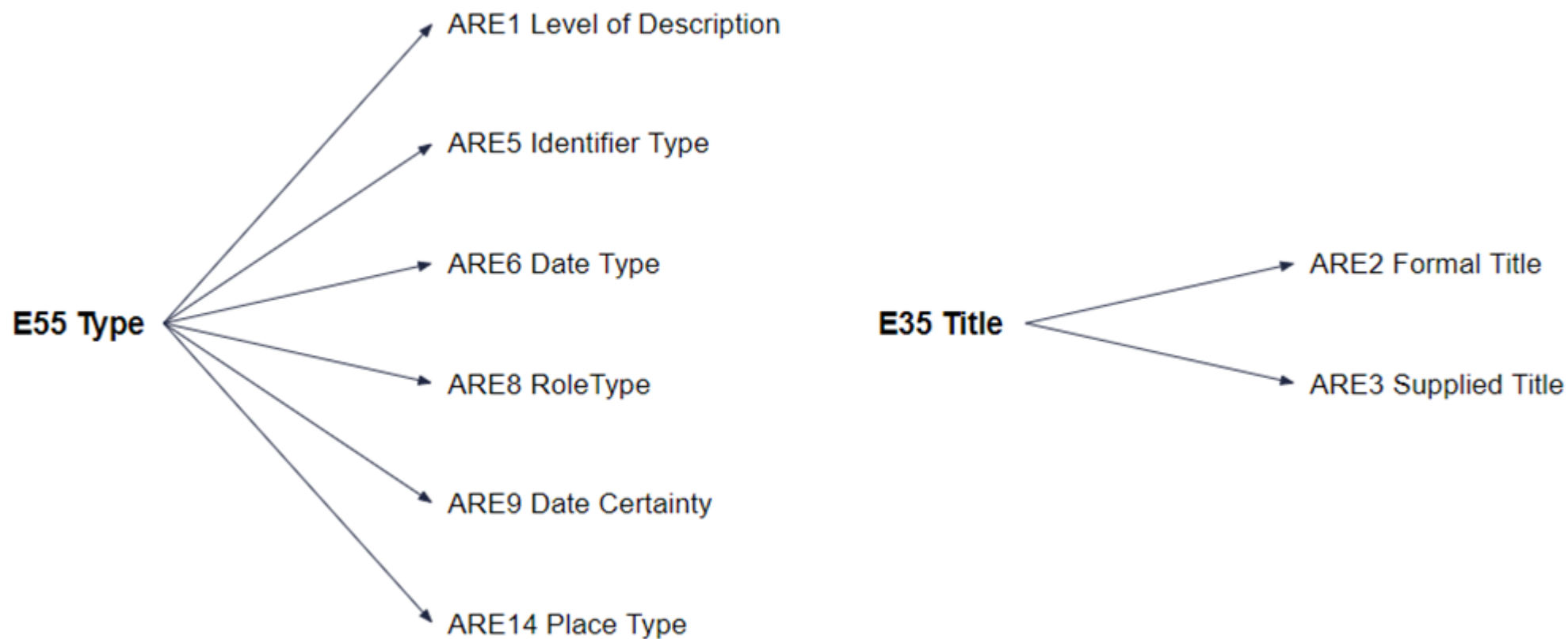| Class | Vocabulary | Example Values |
|-------|-----------|----------------|
| ARE1 Level of Description | Level of description | Fonds; Series; Section; File; Item |
| ARE2 Formal Title<br>ARE3 Supplied Title | Title Type | Formal, Supplied |
| ARE5 Identifier Type | Identifier of collective person/group | PT; VCT; AGH01; 161016; ADLSB; 600084892; PT-LiBN |
| ARE6 Date Type<br>ARE9 Date Certainty | Type of time period | Exact dates; Inferred dates; Predominant dates |
| ARE7 Name Type | Type of name of collective person/group | Authorized form of name; Another form of the name; Parallel name form |
| ARE8 Role Type | Role played | Producer; Material Author; Recipient |
| ARE11 Documentary Typology | Documentary Typology | Certificate; Income book; Patent |
| ARE13 Subject Type | Subject | Education; Science; Law; Management |
| ARE14 Place Type | Type of jurisdictional entity | Ocean; Archipelago; Mountain range; Country; District |
| ARE15 Acquisition Type | Transfer of Custody / Acquisition Identifier | Purchase; Giving; Donation; Deposit; Swap; Legacy; Reintegration; Transfer |
| ARE16 Event Type | Event Type | Evaluation; Expertise; Financial management |
| E56 Language | Language Identifier | Portuguese; Latin; French; Greek |
| E57 Material | Support | Paper; Parchment; Photosensitive film |
| E58 Measurement Unit | Measurement Unit | Centimeter; Gram; Byte; Minute; Pack |
| E98 Currency | Currency | Euro; Dollar; Kwanza |

# ArchOnto

# ArchOnto - CIDOC CRM
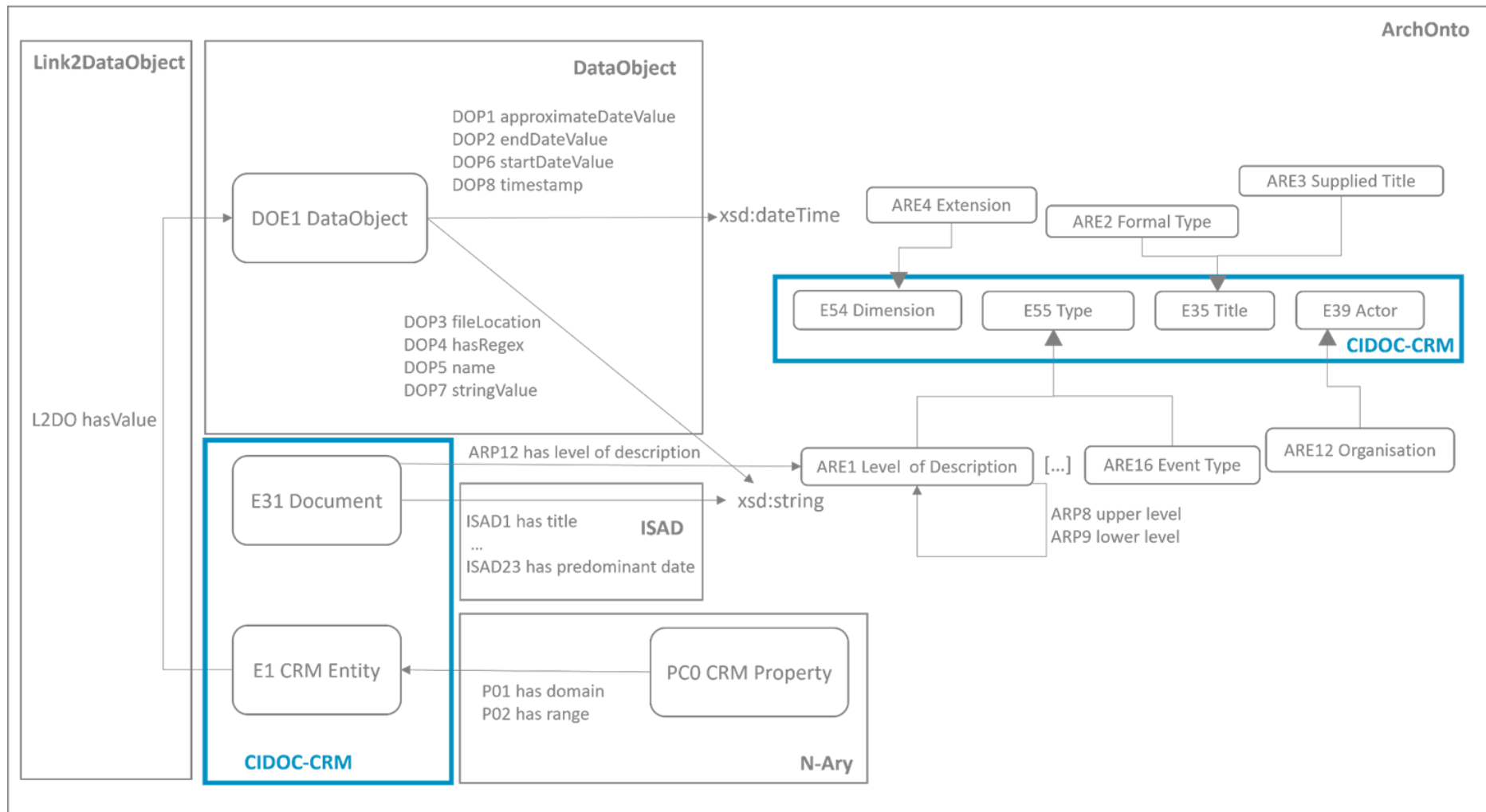
# ArchOnto - DataObject
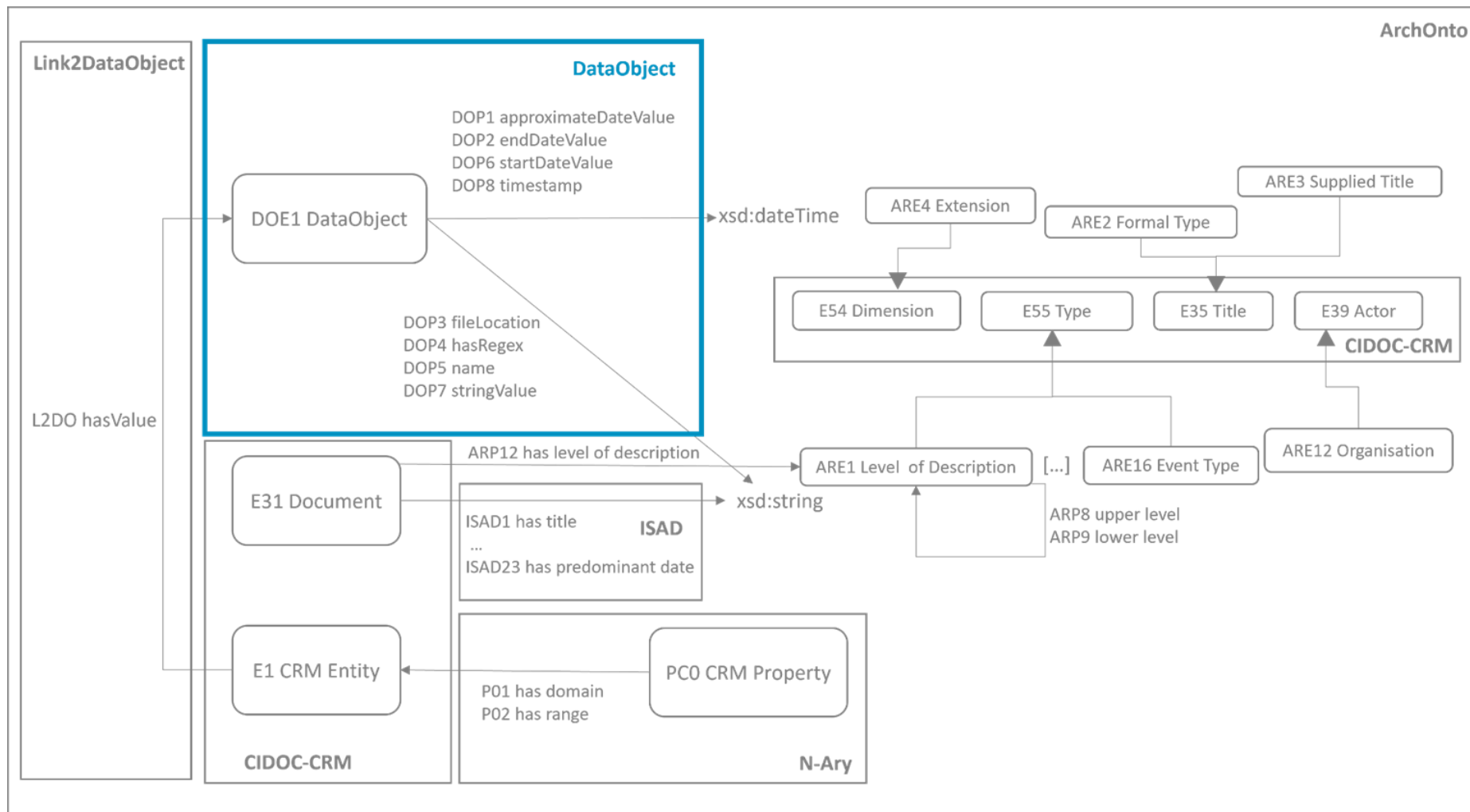
# ArchOnto - N-ary
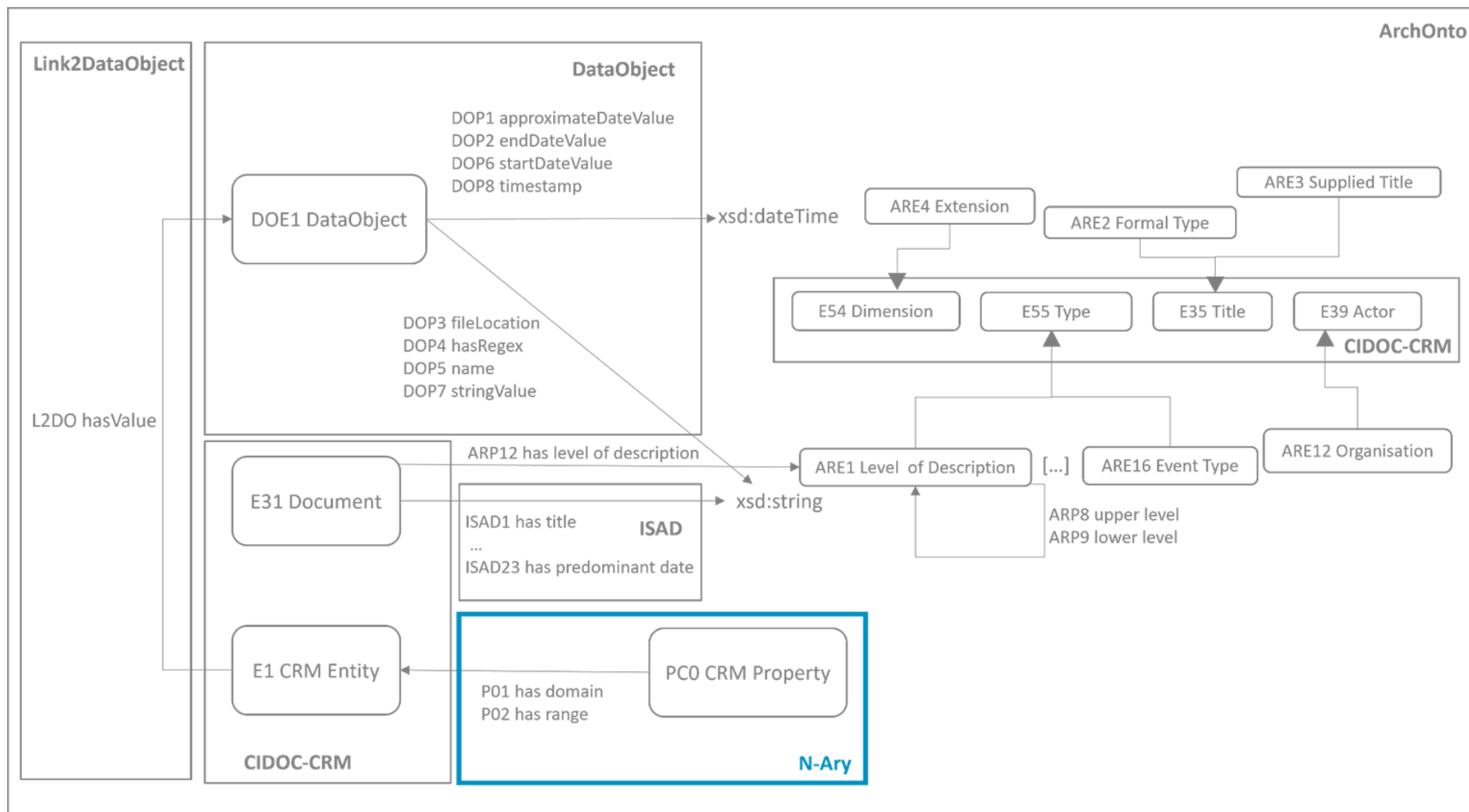
# ArchOnto - ISAD Ontology

# ArchOnto - Link2DataObject

# Categorization of Archival Concepts - ArchOnto



E55 Type
- ARE1 Level of Description
- ARE5 Identifier Type
- ARE6 Date Type
- ARE8 RoleType
- ARE9 Date Certainty
- ARE14 Place Type

E35 Title
- ARE2 Formal Title
- ARE3 Supplied Title

# Archival Records Representation

# DigitArq Record
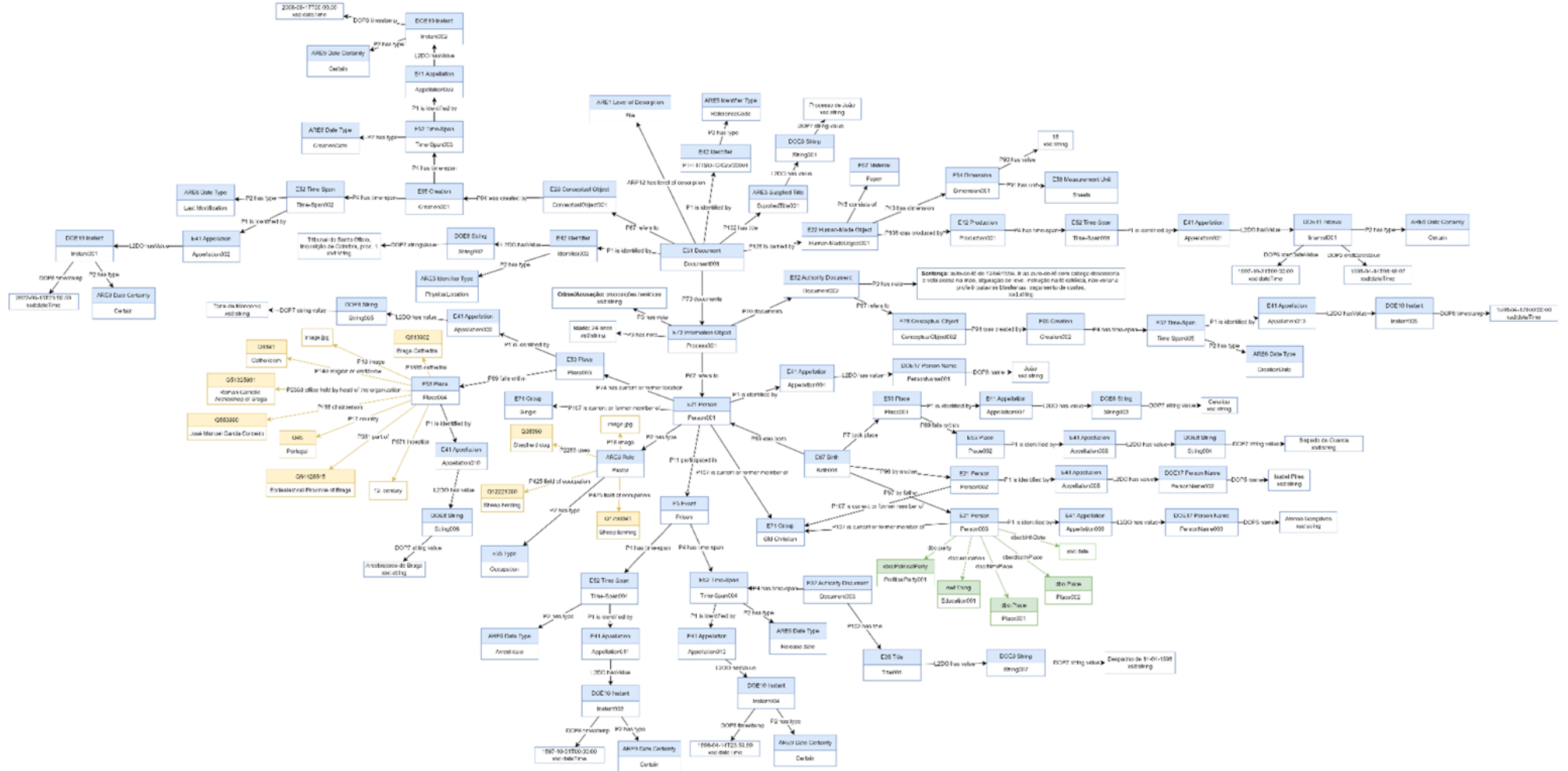
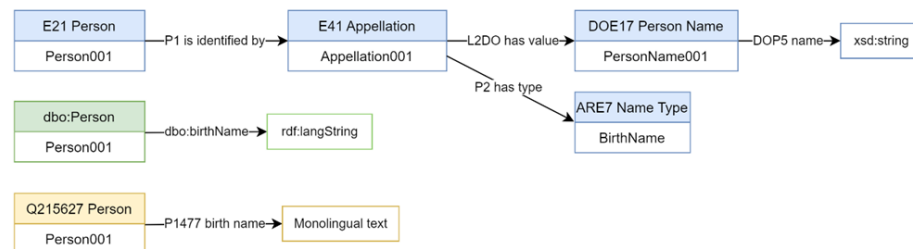# ArchOnto Representation

# DBpedia Representation
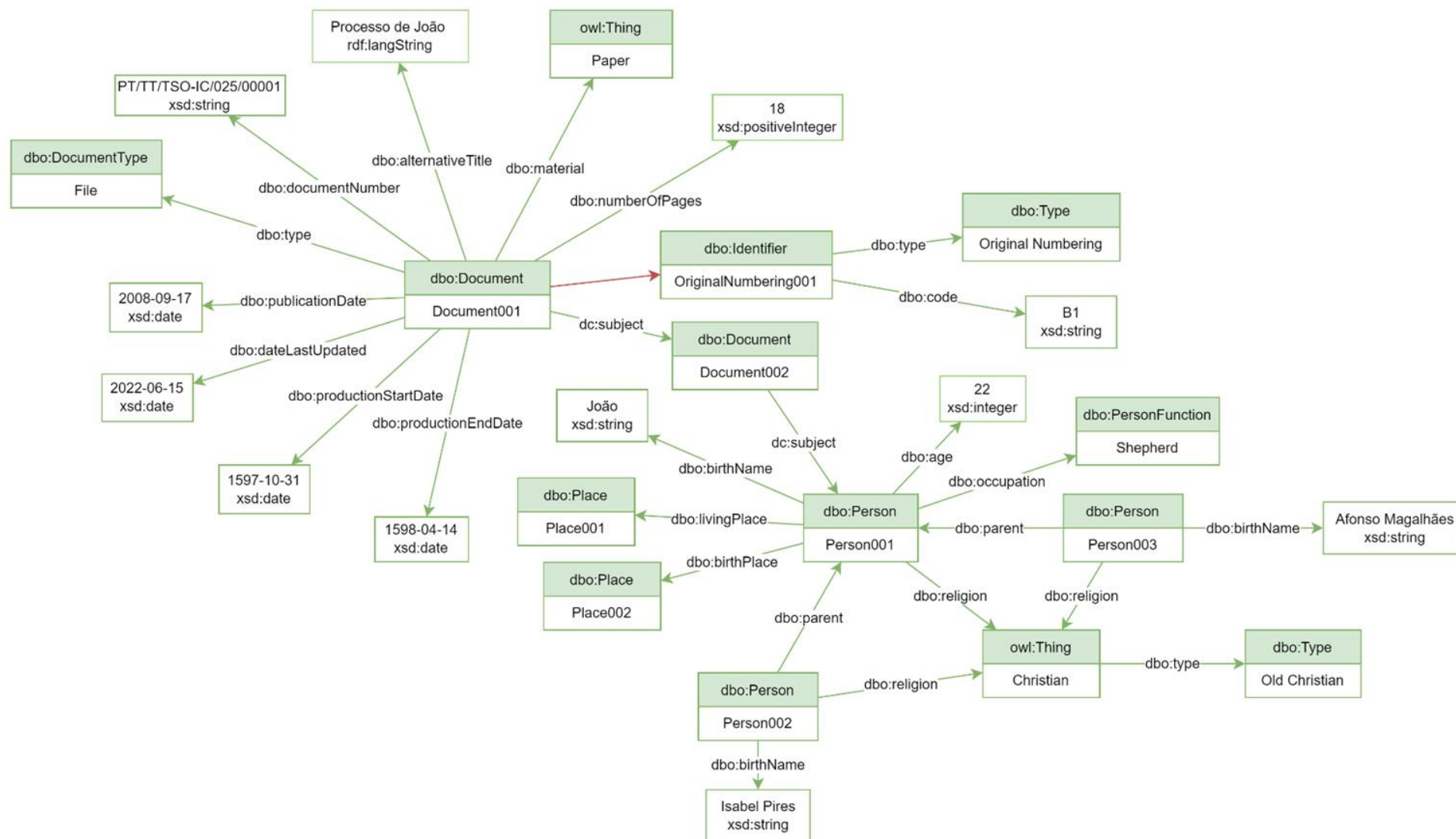
# Wikidata Representation
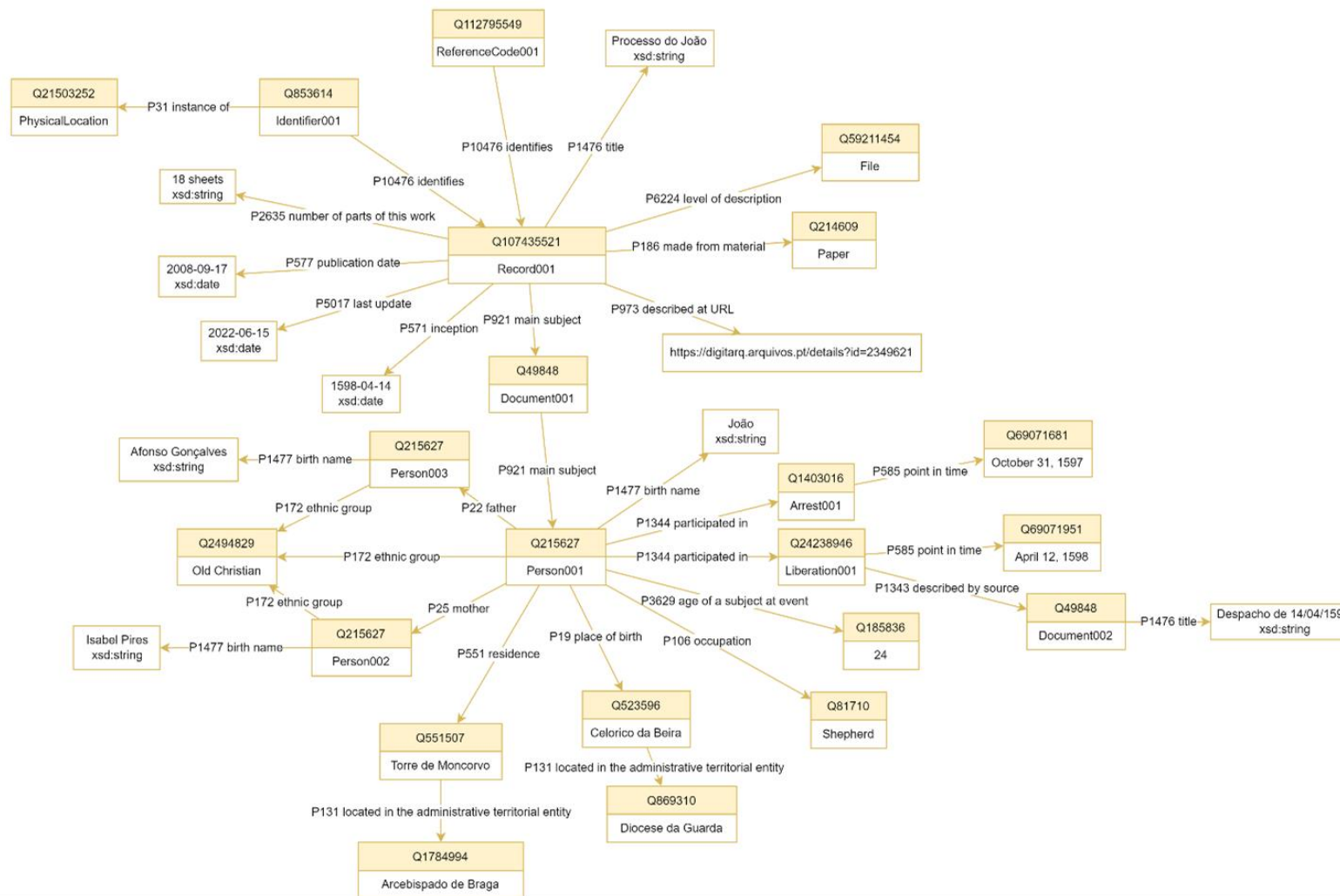
# ArchOnto + DBpedia + Wikidata Representation

# ArchOnto VS DBpedia VS Wikidata
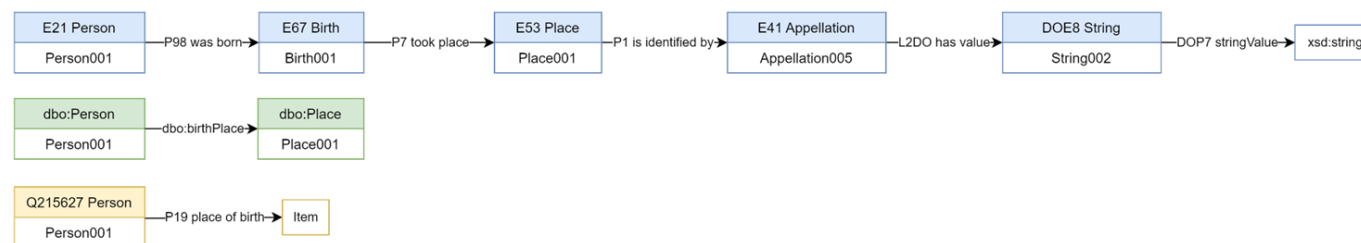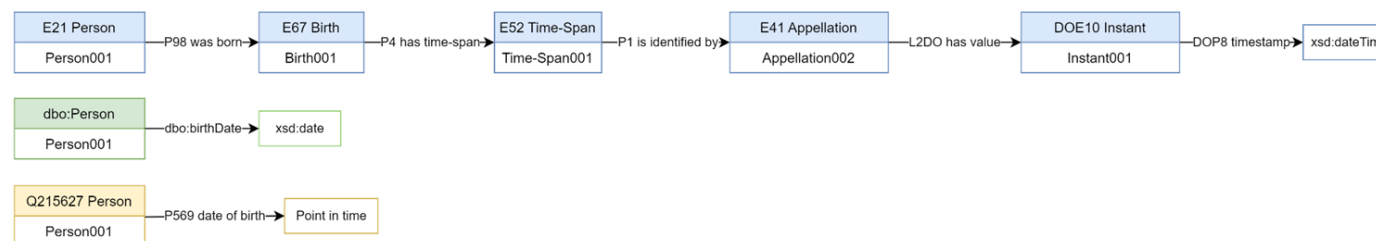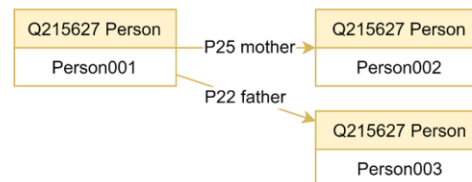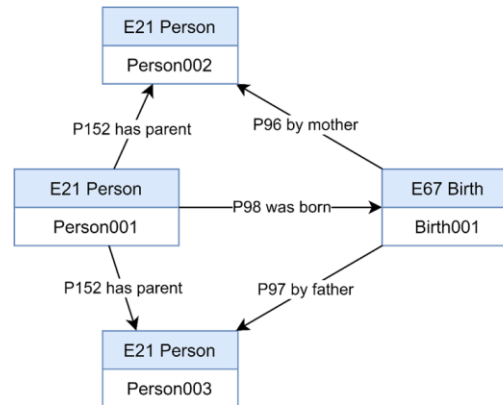
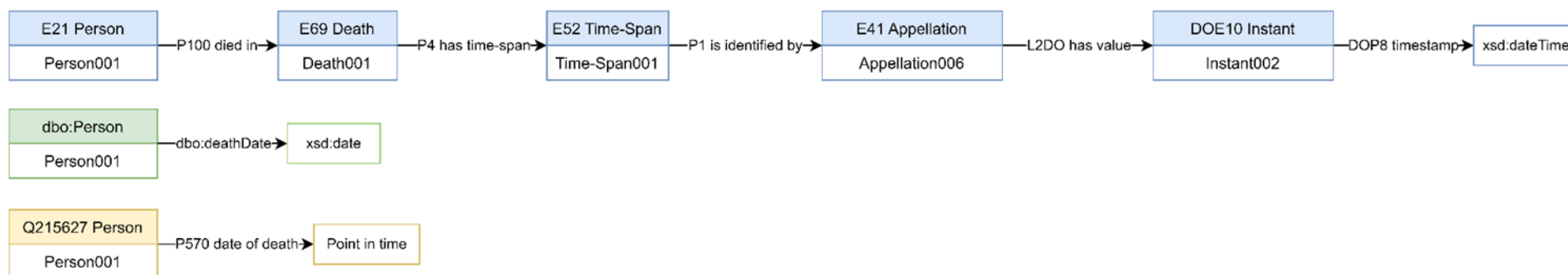# ArchOnto VS DBpedia VS Wikidata



Mother, father and son
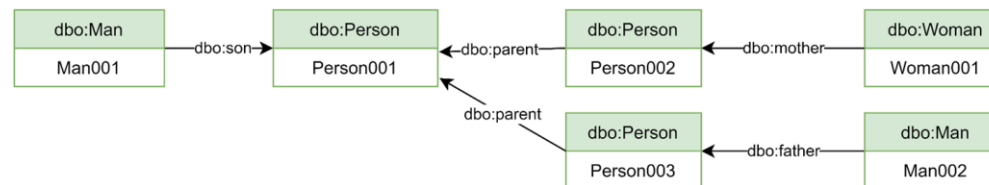
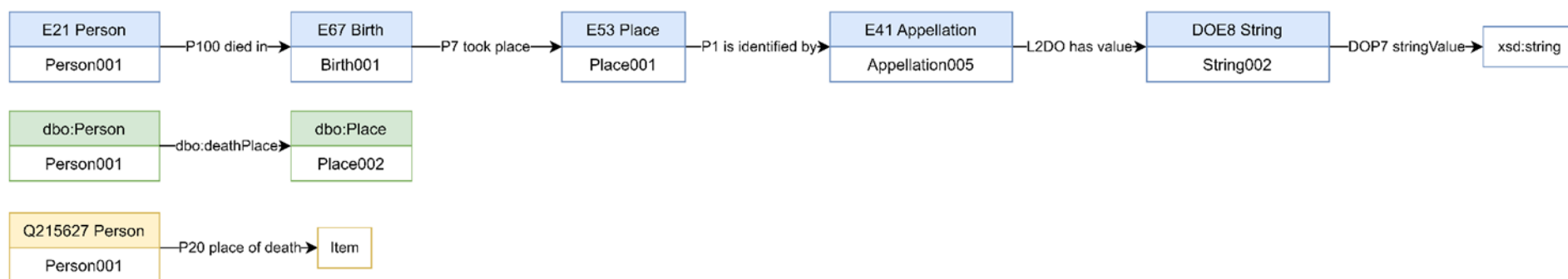# ArchOnto VS DBpedia VS Wikidata

# ArchOnto VS DBpedia VS Wikidata

**Title of a document**

| E31 Document | | ARE3 Supplied Title | | DOE8 String | | |
|---|---|---|---|---|---|---|
| Document001 | —P102 has title→ | SuppliedTitle001 | —L2DO hasValue→ | String001 | —DOP7 stringValue→ | xsd:string |

| E31 Document | | ARE2 Formal Title | | DOE8 String | | |
|---|---|---|---|---|---|---|
| Document001 | —P102 has title→ | FormalTitle001 | —L2DO hasValue→ | String001 | —DOP7 stringValue→ | xsd:string |

| E31 Document | | E35 Title | | DOE8 String | | |
|---|---|---|---|---|---|---|
| Document001 | —P102 has title→ | Title001 | —L2DO hasValue→ | String001 | —DOP7 stringValue→ | xsd:string |

| dbo:Document | | |
|---|---|---|
| Document001 | —dbo:originalTitle→ | rdf:langString |

| dbo:Document | | |
|---|---|---|
| Document001 | —dbo:alternativeTitle→ | rdf:langString |

| Q107435521 Record | | |
|---|---|---|
| Record001 | —P1476 title→ | Monolingual text |

| Q49848 Document | | |
|---|---|---|
| Document001 | —P1476 title→ | Monolingual text |

# 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
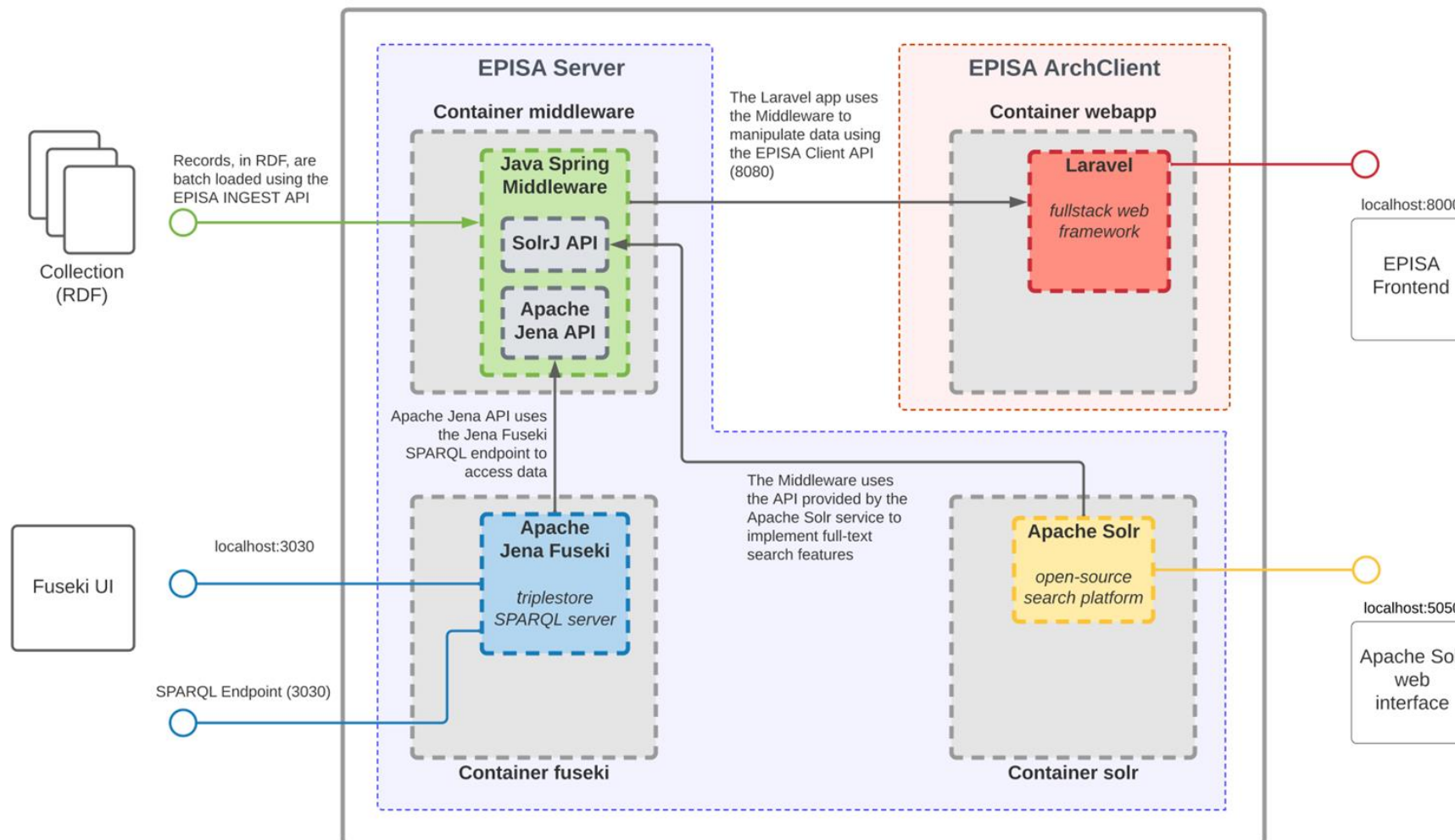
# User Interface Abstractions

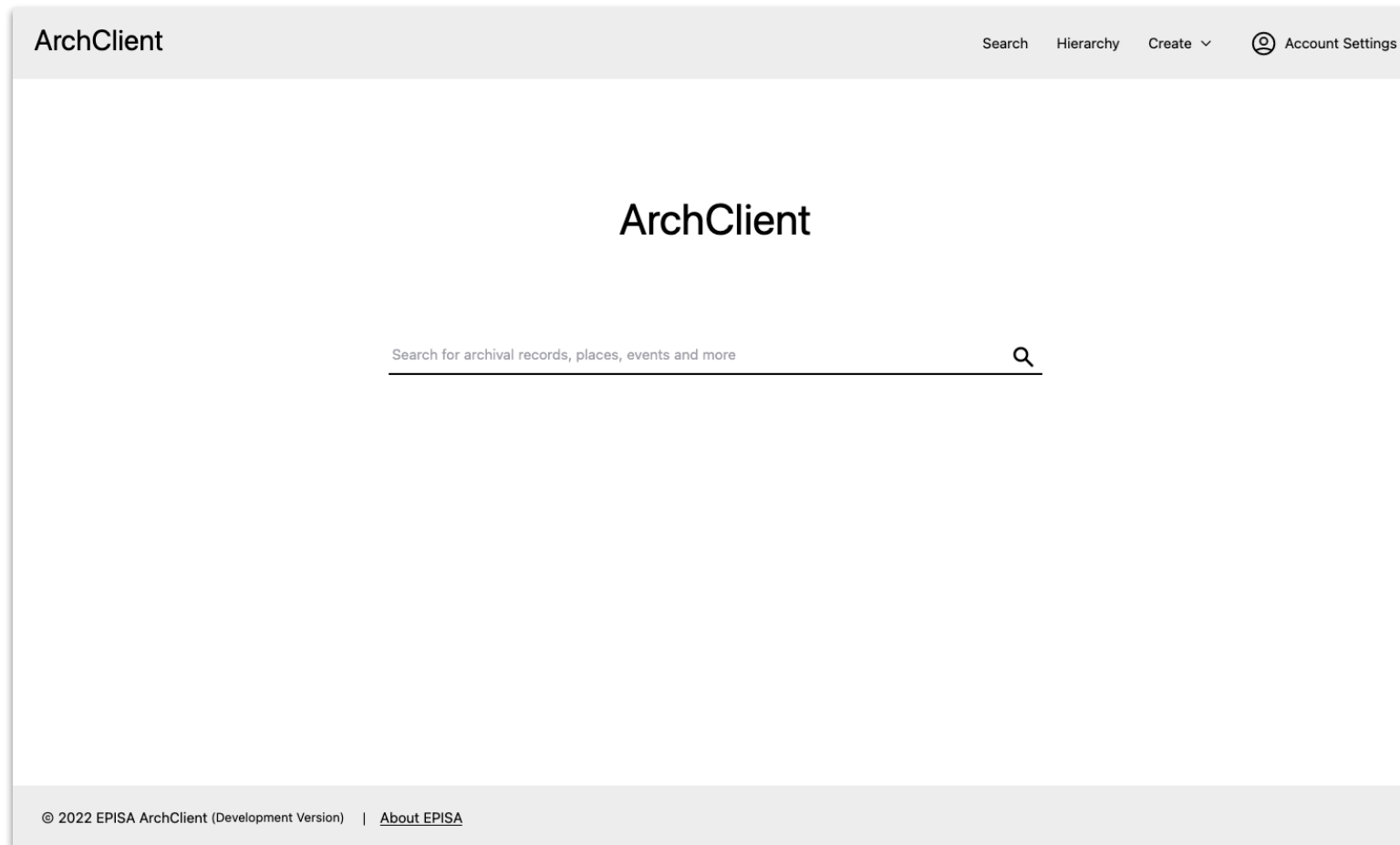Higher-level abstractions need to be defined at the user interface level.

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
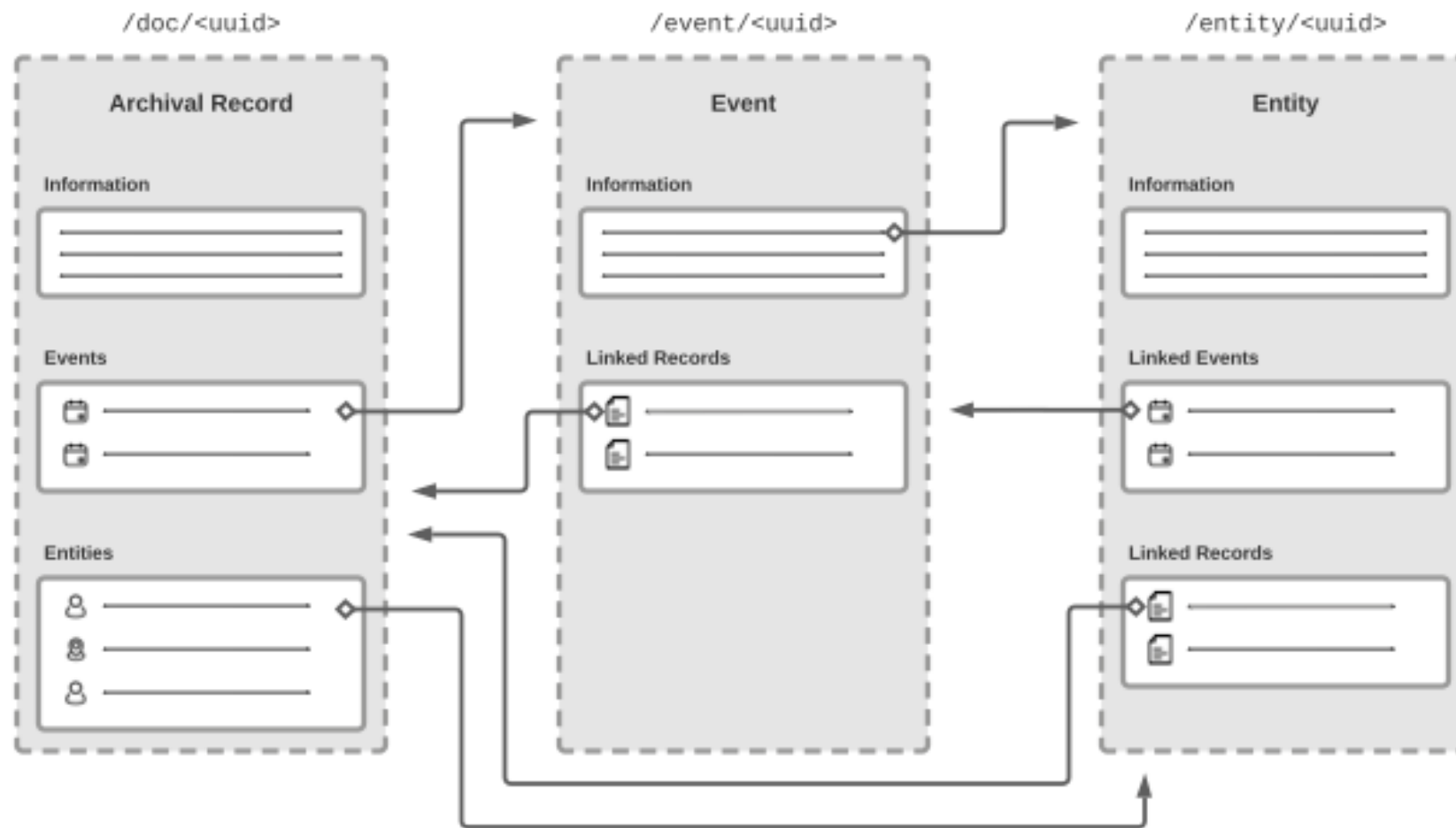
# EPISA ArchClient

# 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

INESCTEC

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Ontology Engineering Group