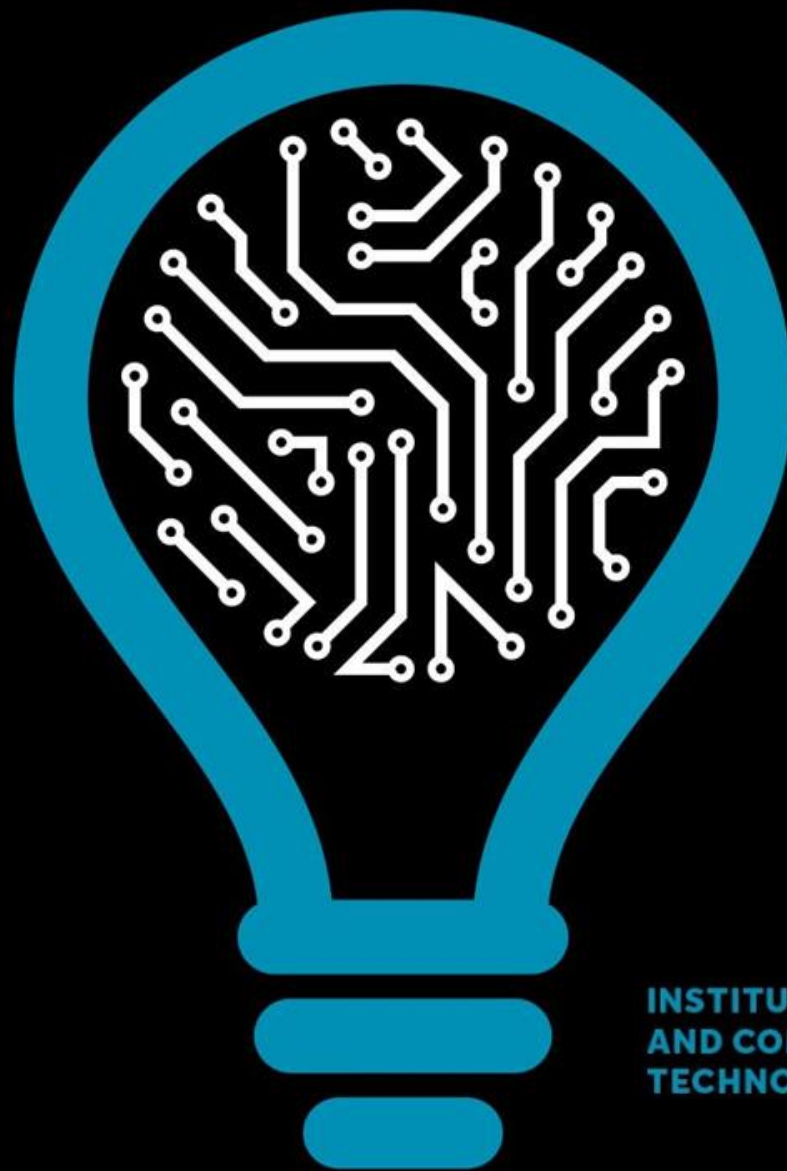


from knowledge  
production to  
science-based  
innovation



**U. PORTO**  
FEUP FACULDADE DE ENGENHARIA  
UNIVERSIDADE DO PORTO

 **INESCTEC**

**INSTITUTE FOR SYSTEMS  
AND COMPUTER ENGINEERING,  
TECHNOLOGY AND SCIENCE**

# ArchOnto, a CIDOC-CRM-based Linked Data Model for the Portuguese Archives

TPDL Conference 2020

**Inês Koch**

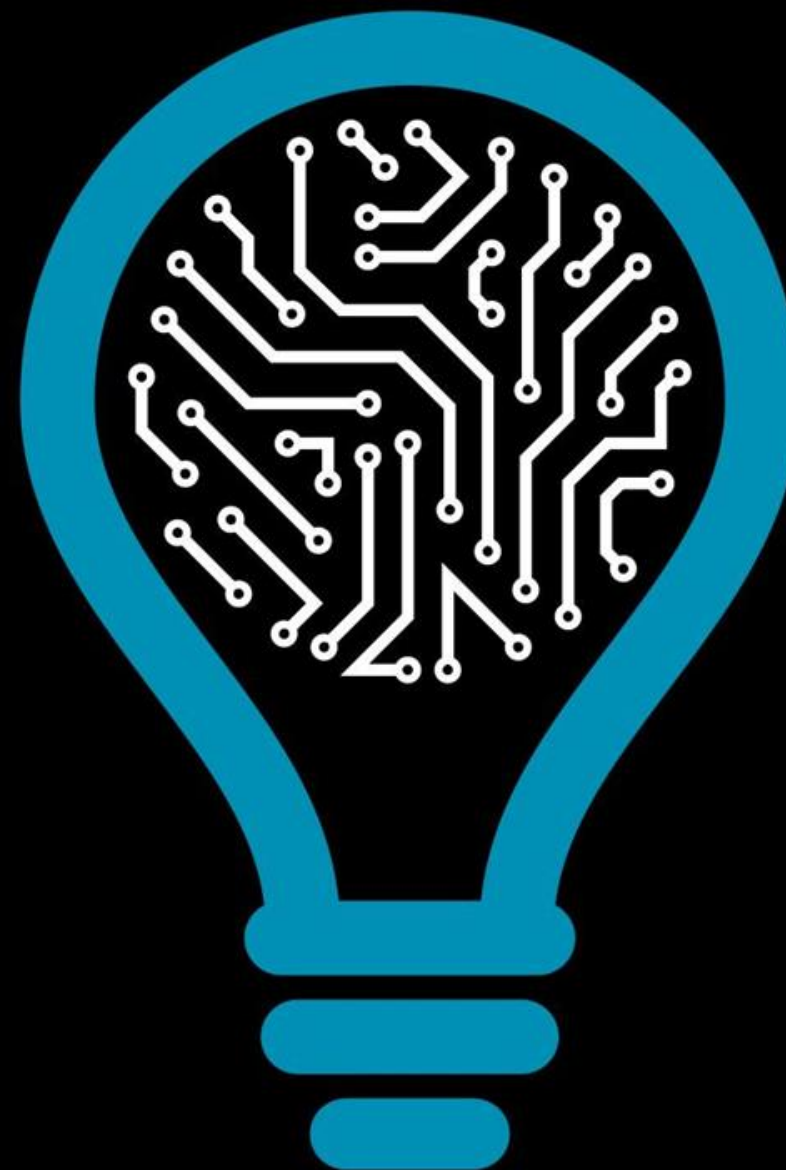
Cristina Ribeiro

Carla Teixeira Lopes

**U. PORTO**  
FEUP FACULDADE DE ENGENHARIA  
UNIVERSIDADE DO PORTO

 **INESCTEC**

INSTITUTE FOR SYSTEMS  
AND COMPUTER ENGINEERING,  
TECHNOLOGY AND SCIENCE



# Outline

- EPISA Project
- Standards for Cultural Heritage
- ArchOnto, a modular ontology for archives
- Evaluation of CIDOC-CRM for archives
- Ongoing work

# EPISA

Entity and Property Inference for  
Semantic Archives

# EPISA team

**INESC TEC:** Cristina Ribeiro, Carla Teixeira Lopes, Gabriel David, João Rocha da Silva, Sérgio Nunes, Bruno Giesteira, Inês Koch, Nuno Freitas, Lázaro Costa, Cláudia Guedes

**U.Évora:** Irene Rodrigues, Dora Simões

**DGLAB:** Francisco Barbedo, Ana Maria Rodrigues, Lucília Runa, Maria José Almeida



**INESCTEC**



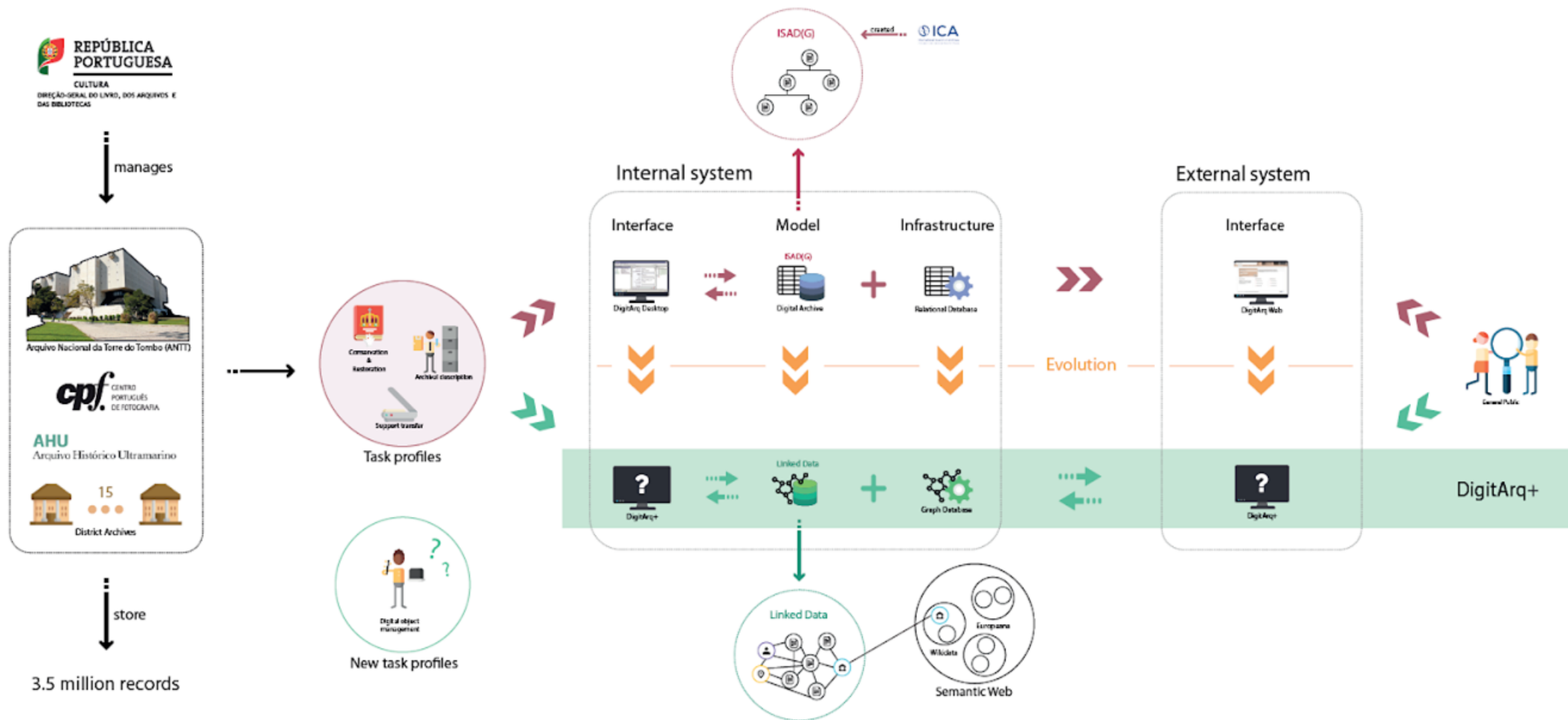
UNIVERSIDADE  
DE ÉVORA



REPÚBLICA  
PORTUGUESA

CULTURA  
DIREÇÃO-GERAL DO LIVRO, DOS ARQUIVOS E  
DAS BIBLIOTECAS

# EPISA



# Standards for Cultural Heritage

ISAD(G), CIDOC-CRM, RiC-CM

# ISAD(G) vs CIDOC-CRM vs RiC-CM

	ISAD(G)	CIDOC-CRM	RiC-CM
Hierarchical	✓	✗	✓
LOD	✗	✓	✓
Ontology	✗	✓	✓
Number of properties	26 elements	285 properties	449 properties
Institutional support	ICA (+25 years)	ICOM (+15 years)	ICA (5 years)
Implementation on archives	Custom	(none known)	None
Most recent version	2000	2020	2020
Support	ICA	CIDOC-CRM SIG	ICA (EGAD)



# ArchOnto

A modular ontology for archives

# Adapting CIDOC-CRM for archival use

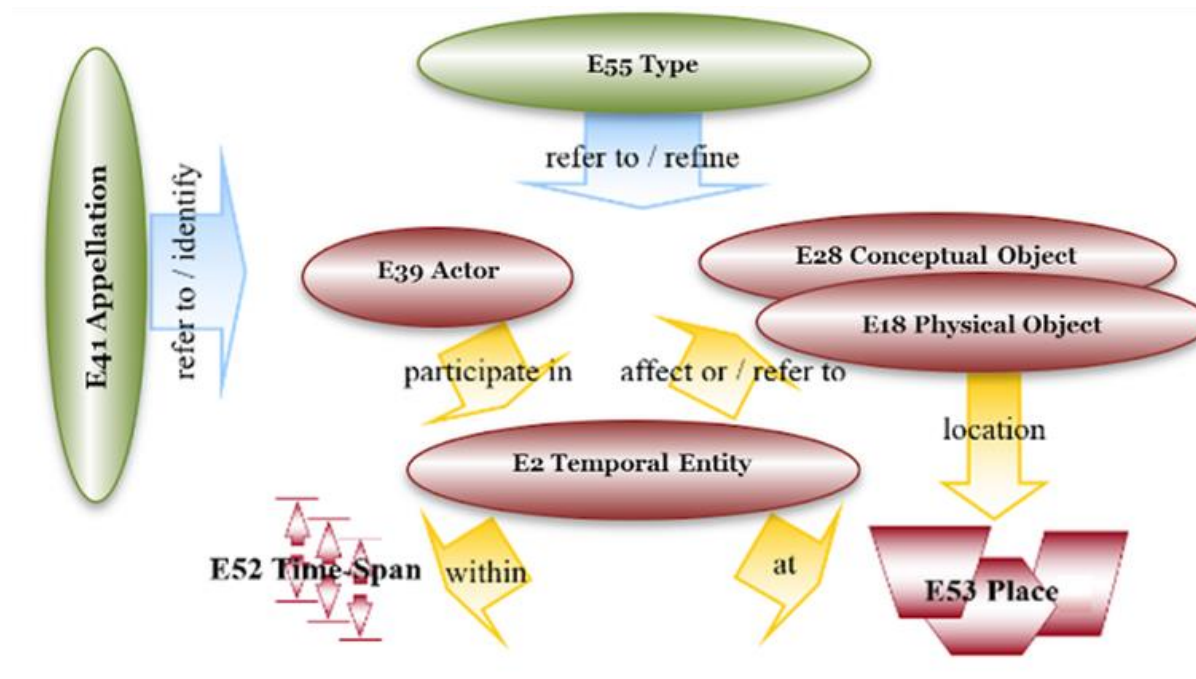
- CIDOC-CRM Extensions
  - Create Data Property extensions to accommodate information from the text fields of the ISAD(G) elements.
- From CIDOC-CRM extensions to separate ontologies
  - Creation of ISAD Ontology and N-ary ontology
- Remaining CIDOC-CRM extensions
  - Classes and Object Properties considered essential to archives, remained in ArchOnto

# Ontologies in ArchOnto

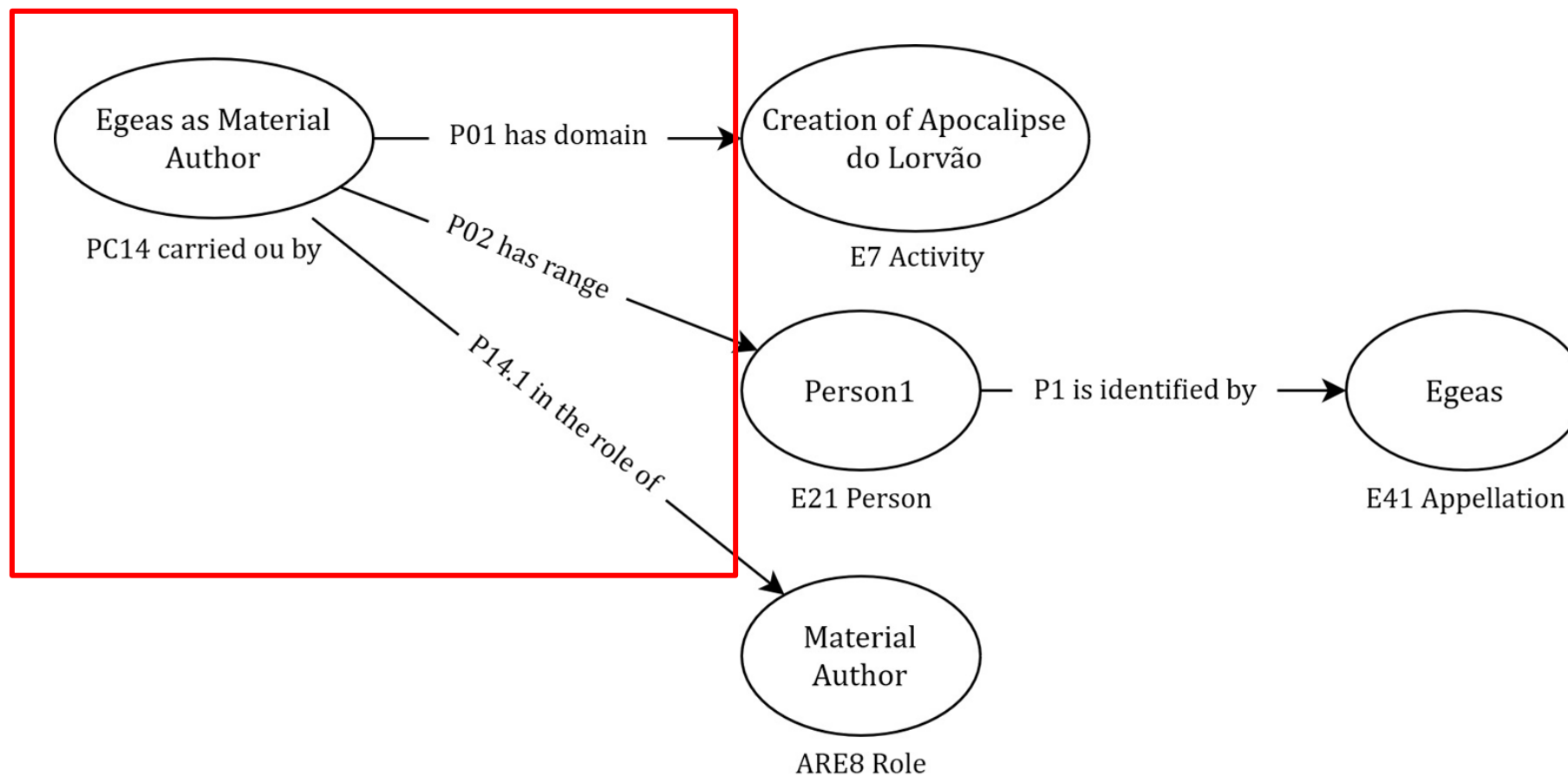
- ArchOnto is composed by 5 different ontologies:
  - CIDOC-CRM
  - N-ary
  - ISAD Ontology
  - DataObject
  - Link between DataObject and CIDOC-CRM

# CIDOC-CRM (Conceptual Reference Model)

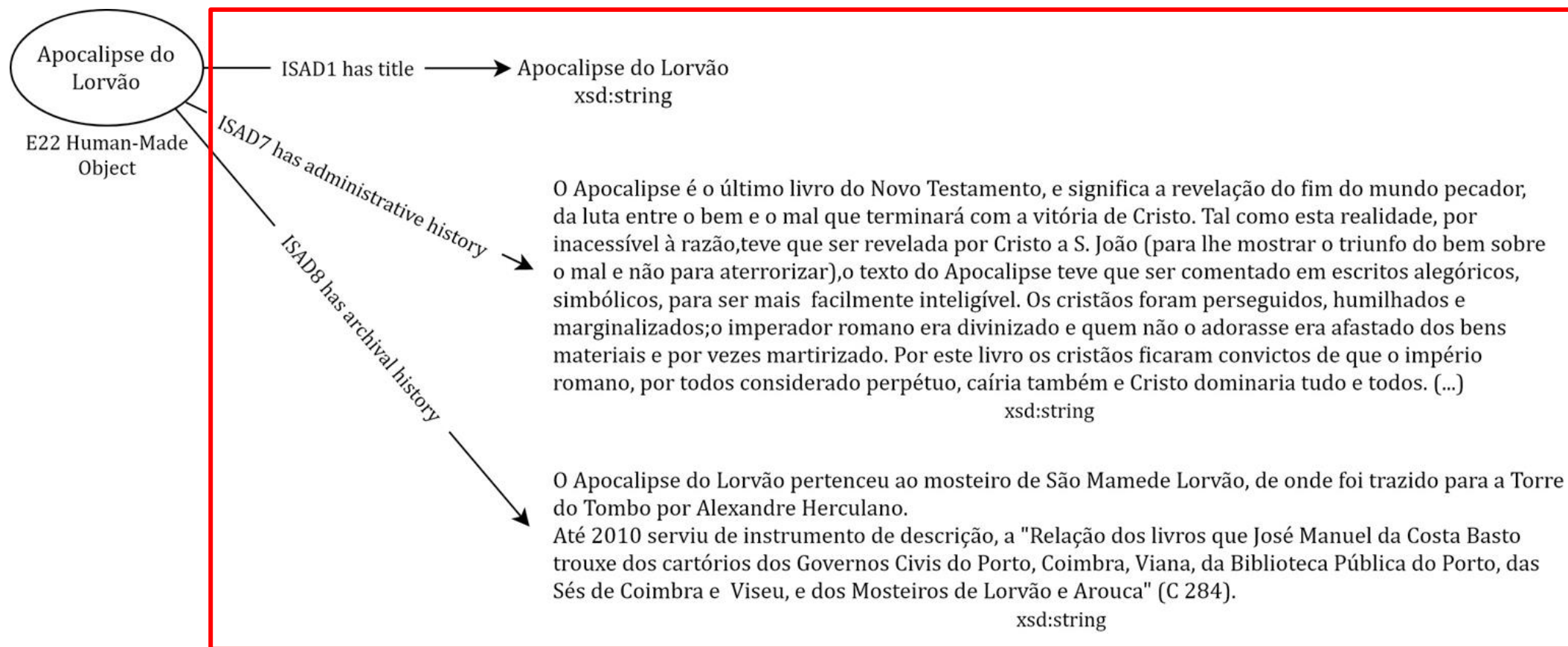
- A formal ontology developed by the International Committee for Documentation (CIDOC) of the International Council of Museums (ICOM)



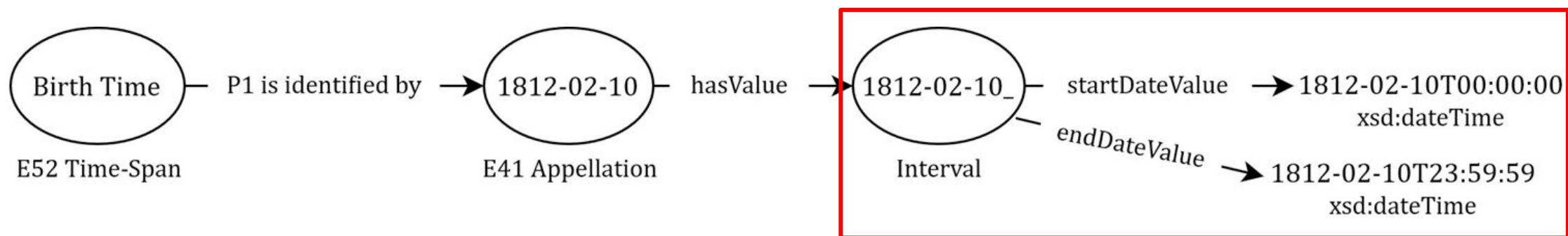
# N-ary



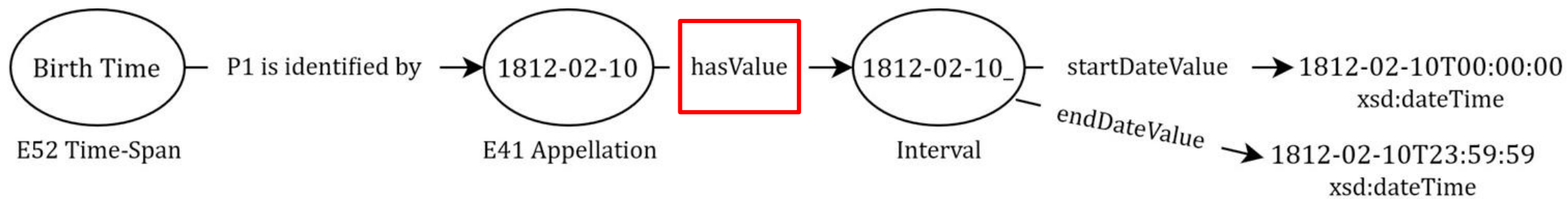
# ISAD Ontology



# DataObject



# Link between *DataObject* and CIDOC-CRM



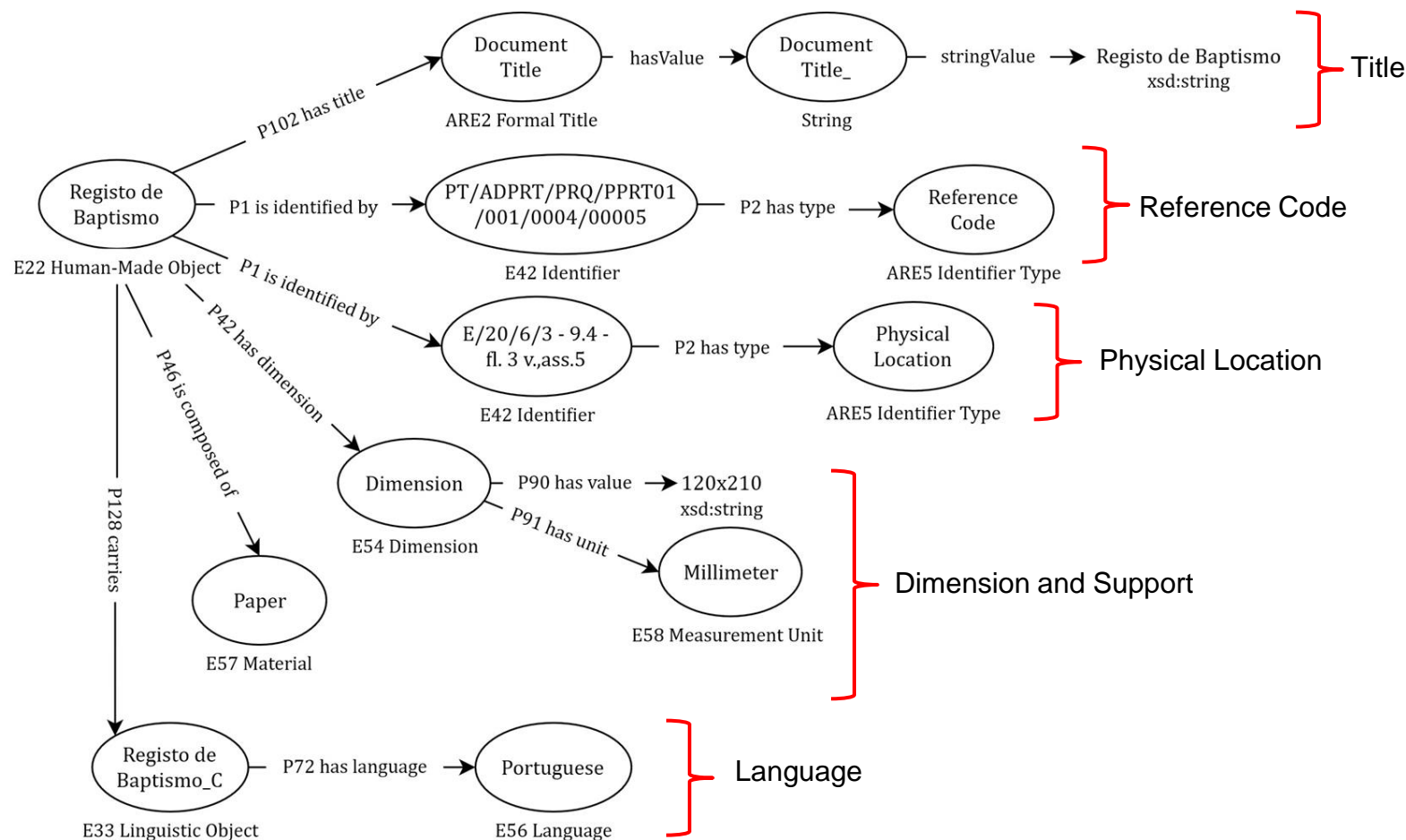


# Issues in the adaptation of CIDOC-CRM for archives

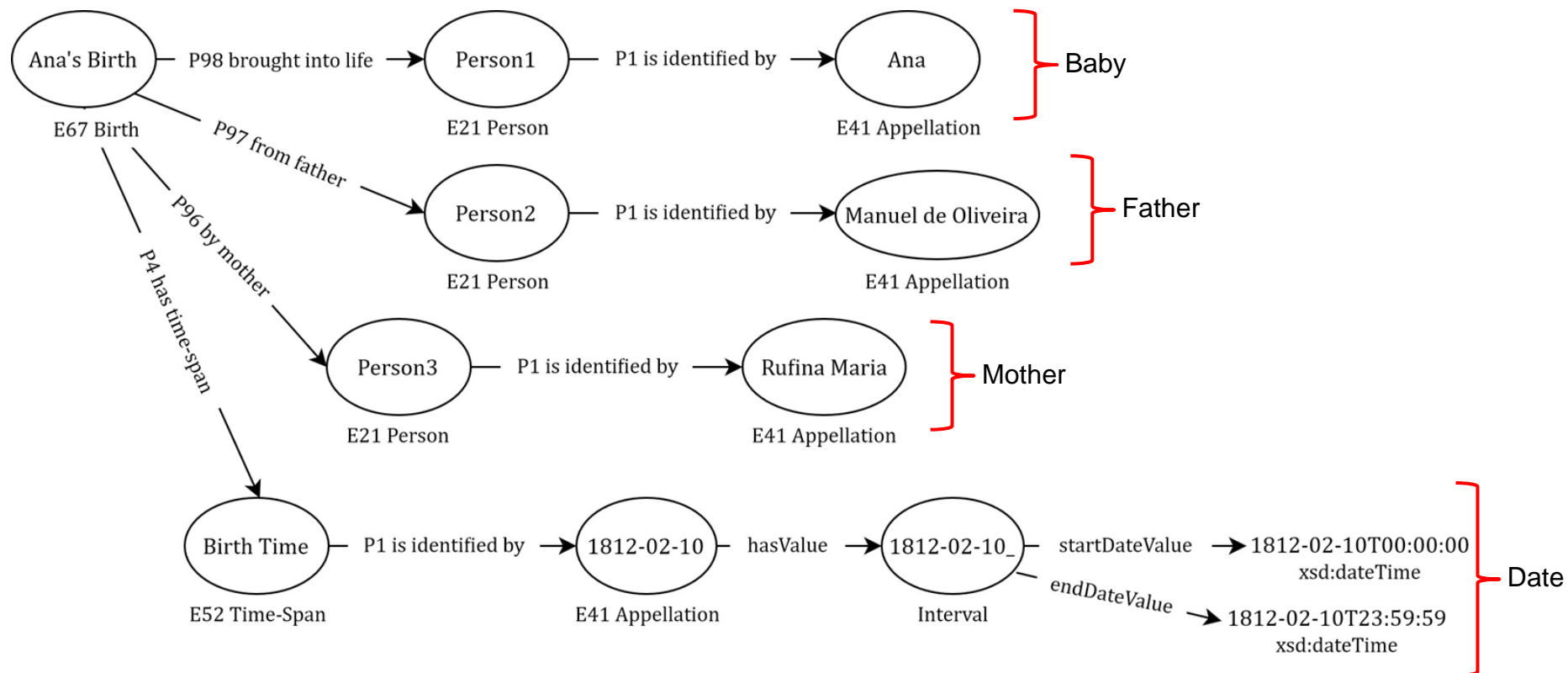
- Types
  - E55 Type class was extensively used, so we decided to create subclasses to have specific types to distinguish them
- Conceptual Object vs Physical Object
  - The need to distinguish physical object from conceptual object appeared, for the first time, with the need to represent a language of a document in CIDOC-CRM
- Validation of data types
  - DataObject handles the transition to the actual representation as values with validated formats.

# Evaluation of CIDOC-CRM for archives

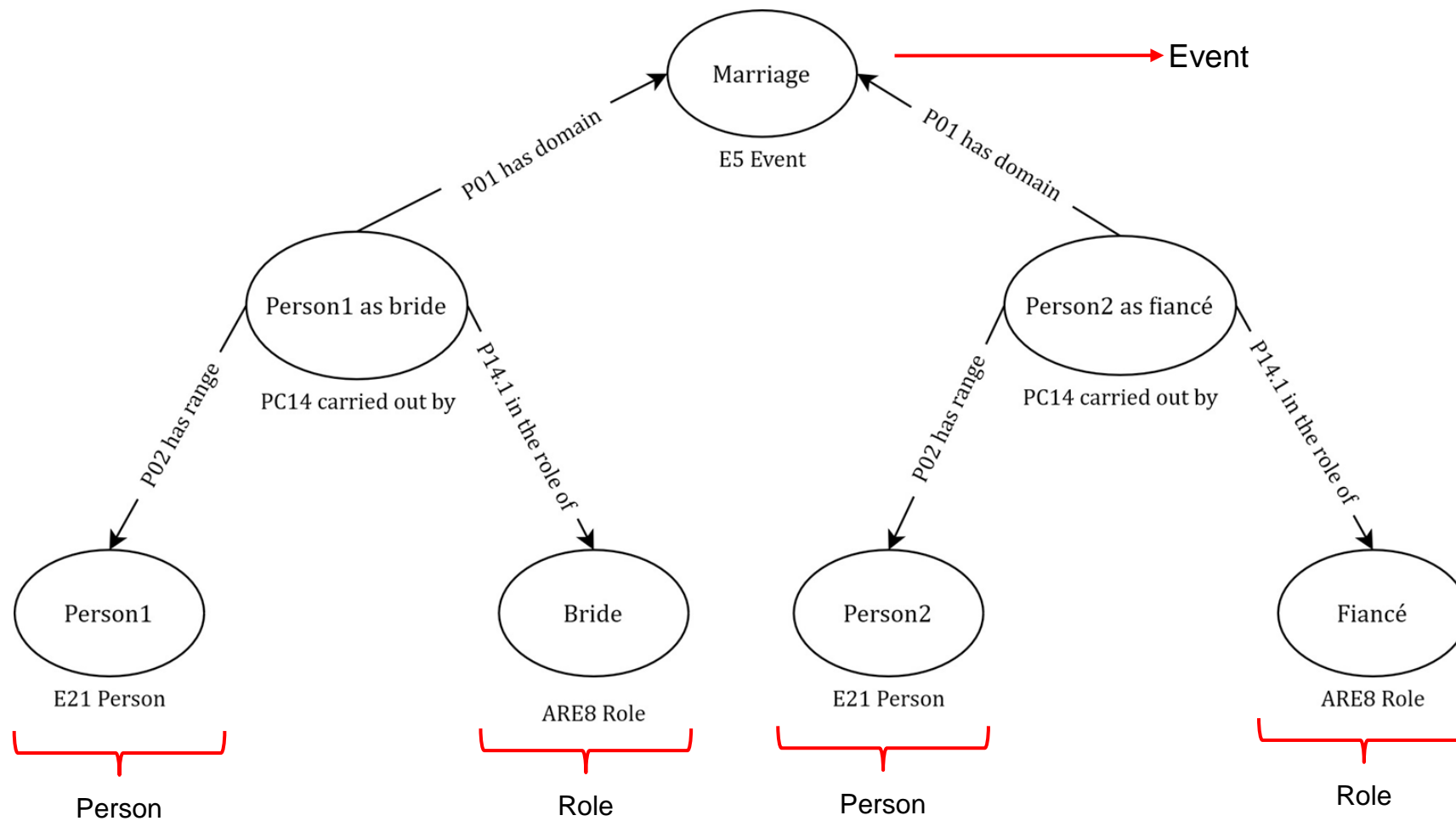
# Evaluation of ISAD(G) concepts in CIDOC-CRM



# Birth event at CIDOC-CRM



# Marriage event at CIDOC-CRM



# Future work

# Future work

- Migration of the ISAD(G) records to the new data model
- Choice of the technology for the knowledge graph
- Prototyping of a system for archives based on Linked Open Data

# ArchOnto, a CIDOC-CRM-based Linked Data Model for the Portuguese Archives

TPDL Conference 2020

Inês Koch, Cristina Ribeiro and Carla Teixeira Lopes

**Follow us:**

<http://episa.inesctec.pt>

<https://github.com/feup-infolab/archontology>

<https://github.com/feup-infolab/archgraph>



INSTITUTE FOR SYSTEMS  
AND COMPUTER ENGINEERING,  
TECHNOLOGY AND SCIENCE

