

# EPISA – Entity and Property Inference for Semantic Archives

Inês Koch, Carla Teixeira Lopes, Cristina Ribeiro, Sérgio Nunes

Faculty of Engineering of the University of Porto, Porto, Portugal

INESC TEC, Porto, Portugal

Envisioning the renewal of the existing Portuguese National Archives data infrastructure, the EPISA Project<sup>1</sup> was created. Besides the National Archives, this project involves the INESC TEC research institute and the University of Évora.

The EPISA has two main goals. The first one consists of the definition of a model with explicit representation for existing archival records to allow the incorporation of the archives in the global network of semantically linked data. The second proposes to develop a prototype for an open-source knowledge graph platform, which will represent archival information in a linked data model. The prototype consists in interfaces that enable a more efficient interaction for specialists, general users, and machines. This prepares the archives to deal with more systematic ingestion of born-digital fonds and metadata interoperability on physical and digital assets.

The developed model is called ArchOnto<sup>2</sup>, a CIDOC CRM extension for archives, where the metadata standards currently used in archives, namely ISAD (G) and ISAAR (CPF), are mapped to combinations of CIDOC CRM classes and properties. ArchOnto has five ontologies at its base, which complement each other. Besides the CIDOC CRM (base ontology), this ontology comprises N-ary<sup>3</sup>, ISAD Ontology<sup>4</sup>, DataObject<sup>5</sup>, and Link2DataObject<sup>6</sup> (Koch, Ribeiro, and Teixeira Lopes 2020).

N-ary provides a systematic way to represent non-binary associations, i.e., those that connect more than two individuals. ISAD Ontology matches the ISAD (G) standard elements that can be directly obtained from existing records. The information is represented with Data Properties and is not atomized. DataObject is the auxiliary ontology to deal with literal values and their validation in ArchOnto and comprises classes and data properties for the simple types in the ontology. Link2DataObject connects the CIDOC CRM and the DataObject ontologies with a single property to make this connection.

---

<sup>1</sup> Website EPISA – Last consulted on 18/10/2022. Available at – <https://episa.inesctec.pt>

<sup>2</sup> Website ArchOnto – Last consulted on 18/10/2022. Available at – <https://purl.org/episa/archonto>

<sup>3</sup> Website N-ary – Last consulted on 18/10/2022. Available at – [https://purl.org/episa/archonto/n\\_ary](https://purl.org/episa/archonto/n_ary)

<sup>4</sup> Website ISAD Ontology – Last consulted on 18/10/2022. Available at – [https://purl.org/episa/archonto/isad\\_ontology](https://purl.org/episa/archonto/isad_ontology)

<sup>5</sup> Website DataObject – Last consulted on 18/10/2022. Available at – [https://purl.org/episa/archonto/data\\_object](https://purl.org/episa/archonto/data_object)

<sup>6</sup> Website Link2DataObject – Last consulted on 18/10/2022. Available at – [https://purl.org/episa/archonto/link2data\\_object](https://purl.org/episa/archonto/link2data_object)

The EPISA Platform is a technical infrastructure designed and developed to support archival records management and access using linked data technologies. The architecture of this platform consists of two main components – the EPISA Server, responsible for storage, reasoning, authorization, and search; and a frontend component, the EPISA ArchClient, responsible for user interaction. The platform adopts a client-server paradigm between these components, at a lower level, it follows a microservice-oriented architecture as both components constitute a group of independent services (Nunes et al. 2022).

In this presentation, we intend to describe the EPISA Project, with a particular focus on ArchOnto and its connection to CIDOC, presenting the prototype and the developed interfaces.

**Key Words:** EPISA, CIDOC CRM, ArchOnto, Archival description

## References

- Koch, Inês, Cristina Ribeiro, and Carla Teixeira Lopes. 2020. “ArchOnto, a CIDOC-CRM-Based Linked Data Model for the Portuguese Archives.” *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 12246 LNCS: 133–46. [https://doi.org/10.1007/978-3-030-54956-5\\_10](https://doi.org/10.1007/978-3-030-54956-5_10).
- Nunes, Sérgio, Tiago Silva, Cláudia Martins, and Rita Peixoto. 2022. “EPISA Platform: A Technical Infrastructure to Support Linked Data in Archival Management.” [http://ceur-ws.org/Vol-3246/11\\_Paper4.pdf](http://ceur-ws.org/Vol-3246/11_Paper4.pdf).

*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.*