

A practical implementation of the CIDOC CRM
at the Portuguese National Archives

Inês Koch – ines.koch@inesctec.pt

- INESC TEC – Campus da Faculdade de Engenharia da Universidade do Porto -
Rua Dr. Roberto Frias 4200-465, Porto, Portugal

Maria José de Almeida – m-jose.almeida@dglab.gov.pt

- DGLAB, Alameda da Universidade, Edifício da Torre do Tombo, Lisboa,
Portugal

João Rocha da Silva – joaorosilva@gmail.com

- INESC TEC – Campus da Faculdade de Engenharia da Universidade do Porto -
Rua Dr. Roberto Frias 4200-465, Porto, Portugal

The EPISA project (Entity and Property Inference for Semantic Archives) is part of the ongoing renewal of the The Portuguese National Archives' existing data infrastructure. The goal is to develop a prototype for an open-source knowledge graph platform representing archival information on a linked data model. The need for more fine-grained, more flexible and specially more machine-actionable description tools led to the choice of CIDOC-CRM, the ontology originating in museums, as the core ontology.

ArchOnto is an extension of the CIDOC-CRM for archives, where the metadata standards currently used in the archives, namely ISAD(G) and ISAAR(CPF), are mapped to combinations of CIDOC-CRM classes and properties. *ArchOnto* also includes extensions required for the representation of the archival objects and incorporates, as satellite ontologies, *N-ary* (a solution for the representation of non-binary relationships proposed by CIDOC), *DataObject* (for data validation and data properties that CRM does not provide), *ISAD* (to incorporate textual information from ISAD(G)), and an ontology for the link between *DataObject* and CIDOC-CRM.

The design of ArchOnto is also informed by several non-functional requirements of a working prototype. Maintaining the consistency of the graph is crucial, even when concurrent modifications of the underlying graph are performed. At the same time, one cannot rely on an inference engine, as the response times of the system would increase drastically — we briefly discuss some of these constraints and how we modeled ArchOnto to cope with them.

Key Words: CIDOC-CRM, ISAD(G), Archival description, EPISA