

Comparative insights into semantic archival modelling: evaluating RiC-O and ArchOnto representation capabilities

Journal of
Documentation

1003

Lucia Giagnolini

*Department of Classical Philology and Italian Studies, University of Bologna,
Bologna, Italy*

Inês Koch

*Faculty of Engineering, University of Porto, Porto, Portugal and
INESC TEC, Porto, Portugal*

Francesca Tomasi

*Department of Classical Philology and Italian Studies, University of Bologna,
Bologna, Italy, and*

Carla Teixeira Lopes

*Faculty of Engineering, University of Porto, Porto, Portugal and
INESC TEC, Porto, Portugal*

Received 24 December 2024

Revised 22 April 2025

Accepted 26 April 2025

Abstract

Purpose – This study aims to comparatively evaluate two semantic models, ArchOnto (CIDOC CRM based) and Records in Contexts Ontology (RiC-O), for archival representation within the Linked Open Data framework. The research seeks to critically analyse their ability to represent archival documents, events, activities, and provenance through the application on a case study of historical baptism records.

Design/methodology/approach – The study adopted a comparative approach, utilising the two models to represent a dataset of baptism records from a Portuguese parish spanning several centuries. This involved information extraction and conversion processes, transforming XML EAD finding aids into RDF to facilitate more explicit semantic representation and analysis.

Findings – The analysis revealed distinctive strengths and limitations of each semantic model, providing nuanced insights into their respective capacities for archival description. The findings guide cultural heritage institutions in selecting and implementing the most suitable semantic model for their needs and pave the way for semantic alignment between the two models.

Research limitations/implications – Although the case study explored the representation of a wide range of features, potential limitations include the specific contextual constraints of parish records and the need for broader comparative studies across diverse archival contexts.

Originality/value – This paper offers original insights into semantic modelling for archival representations by providing a detailed comparative analysis of two ontological approaches. It offers valuable perspectives for

© Lucia Giagnolini, Inês Koch, Francesca Tomasi and Carla Teixeira Lopes. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at [Link to the terms of the CC BY 4.0 licence](#)

Lucia Giagnolini is funded by the European Commission, NextGenerationEU, Investment I.4.1, PNRR Scholarships for Cultural Heritage, Ministerial Decree No. 351 of April 9, 2022.

Inês Koch is financed by National Funds through the Portuguese funding agency, Fundação para a Ciência e a Tecnologia (FCT), within the research grant 2020.08755.BD, with DOI:10.54499/2020.08755.BD.

L. Giagnolini is responsible for sections 2, 3.1, 3.3 and 3.5; I. Koch is responsible for sections 1, 3.2 and 3.4.; C. Teixeira Lopes and F. Tomasi supervised the project, reviewed the paper and contributed to draw its conclusions.



Journal of Documentation

Vol. 81 No. 4, 2025

pp. 1003-1031

Emerald Publishing Limited

e-ISSN: 1758-7379

p-ISSN: 0022-0418

DOI 10.1108/JD-12-2024-0310

1. Introduction

Galleries, Libraries, Archives, and Museums (GLAM) institutions have long embraced the Linked Open Data (LOD) paradigm, diving into a process requiring systematic rethinking of cultural heritage description and cataloguing practices. This process involves deconstructing and reconstructing the typology, granularity, and precision of the data to move beyond document-centric description schemes towards data-centric approaches that emphasise contextual relationships (Daquino, 2021).

The international museums and libraries communities began their efforts more than 2 decades ago, leading to the development of the CIDOC Conceptual Reference Model (CIDOC CRM) (ICOM and CIDOC CRM SIG, 2024) in the museum sector and IFLA Library Reference Model object-oriented (LRMoo) (IFLA LRMoo Working Group and CIDOC CRM SIG 2024) in the bibliographic one, representing the latest stage in a series of conceptual developments originating from the Functional Requirements for Bibliographic Records (FRBR) model (Riva *et al.*, 2023).

As for the archival domain, the International Council on Archives (ICA) began to lay the foundations for Linked Data in 2012, when the ICA Programme Commission (PCOM) established the Expert Group on Archival Description (EGAD) with the mission of developing a comprehensive records description standard that integrates and enhances the existing standards – ISAD(G), ISAAR (CPF), ISDF, and ISDIAH (Clavaud and Wildi, 2021). The goal is being achieved with the development of Records in Contexts (RiC), which addresses the activity of describing records through four complementary parts: Records in Contexts-Foundations of Archival Description (RiC-FAD), Records in Contexts-Conceptual Model (RiCCM), Records in Contexts-Ontology (RiC-O) and Records in Contexts-Application Guidelines (RiC-AG). After years of development, through open dialogue with peers and institutions of the archival sector, version 1.0 of the first three parts was released in late 2023, while work on the fourth part began in early 2024 (International Council on Archives Expert Group on Archival Description, 2023b).

In the same period, parallel to the development of RiC, another initiative stood out in the archival context in relation to LOD, namely the Entity and Property Inference for Semantic Archives (EPISA) project (Institute for Systems and Computer Engineering of Porto, 2024). The EPISA project started in 2019 as a joint effort of a team of Information and Computer Science and archival experts from the Portuguese National Archives with the primary goal of incorporating the Portuguese National Archives into the global LOD network (Koch *et al.*, 2020). In light of the landscape of ontologies applied to the cultural heritage domain and the status of RiC (at that time still preliminary), the EPISA project developed ArchOnto, a model for describing archival resources based on CIDOC CRM and additional ontologies (Koch *et al.*, 2023a).

Given this context, we conducted a comparative analysis of how the two approaches – ArchOnto and RiC-O – allow us to represent archival records and their contexts [10]. Several approaches have been proposed to evaluate and compare semantic models, either automated, semi-automated or manual (Liu *et al.*, 2021; Rudwan and Fonou-Dombeu, 2025; Amini *et al.*, 2025). This study starts from a specific case study, manually assessing how the two models can represent the same dataset. This allowed us to analyse their strengths and limitations in a practical scenario, highlighting differences in expressiveness, structure, and interpretability.

This paper aims to serve two primary purposes. First, it assists institutions that are considering migrating their finding aids to linked data and are seeking to understand which model – RiC-O or ArchOnto (thus CIDOC CRM) – is best suited to their needs. These institutions, whether small archives or large cultural heritage organisations, will benefit from the insights provided here on the features of each model. Second, the paper takes a first step toward harmonising RiC-O and CIDOC CRM by highlighting the potential for alignment between these models.

2. Case study and methods

To perform the comparative analysis, we selected the archival description of the “Registos de Baptismos” of the Parish of Aldoar ([Arquivo Distrital do Porto, 2015](#)), provided by the District Archive of Porto ([Arquivo Distrital do Porto, 2013](#)), as our case study.

“Registos de Baptismos” is a series of the “Paróquia de Aldoar” fonds (with the reference code *PT/ADPRT/PRQ/PPRT01/001*), which consists of 17 books containing records of baptisms performed in the Parish of Aldoar from February 1,644 to 30 March 1911. In the series, each book has been identified as an archival unit, and the description goes down to the individual baptism record, which is identified as an item.

The fonds have been described according to the ISAD(G) standard and are available on the District Archive of Porto’s website ([Arquivo Distrital do Porto, 2022](#)). For research purposes, in light of the need for data conversion, the Archive provided us with a copy of the description in XML EAD format.

The case study was well suited to our objective because it has already been represented in ArchOnto as part of the EPISA project ([Koch et al., 2023a](#); [Pires et al., 2023](#)). This prior representation allowed us to leverage the existing work and focus our efforts on RiC-O, streamlining our research process. Additionally, building on a previously initiated project provided continuity and enabled us to compare the two models’ performance on a well-documented dataset and assess improvements in the ArchOnto model. This continuity facilitated a more comprehensive evaluation, as we could draw on the insights and experiences gained from the initial ArchOnto implementation.

Furthermore, the nature of the data allows for a clear demonstration of the rich network of people, events, and activities present in archival records. Baptism records contain information about the record and the series it belongs to, as well as data about the baptised child, their parents, grandparents, godparents, and the dates of birth and baptism. Consequently, these records provide details on the birth event, the baptism activity, and the individuals who participated in these events. This complexity provides an excellent opportunity to evaluate how well each model handles interconnected data and contextual relationships. It offers insights into their effectiveness in representing individual records and their broader archival context.

Given these considerations, the activities were developed in three stages: scenarios identification and representation, data extraction and semantic enrichment, and XML to RDF conversion (see [Figure 1](#) for an overview).

2.1 Scenarios identification and representation

In the first stage of our work, we began by thoroughly examining the dataset to identify the most interesting scenarios in terms of representation.

We identified four specific scenarios – the representation of the archival document, the birth event, the baptism activity, and the attestation of data provenance. These four aspects enabled a comprehensive representation of individuals, their relationships, the events and activities they participated in, and the broader archival context.

We depicted the four scenarios through semantic maps based on RiC-O and ArchOnto so that they were represented with both models, possibly even with multiple representation options [1].

2.2 Data extraction and semantic enrichment

A comprehensive analysis of the original EAD XML file, which details the baptism records of the parish of Aldoar, was conducted during the second stage of this study (see [Listing 1](#) for an EAD



Figure 1. Stages of the research. Source: Authors’ own work

XML example). This involved extracting and isolating pertinent information from descriptive text strings, such as the child's name, birth date, parents' names, grandparents, godmother and godfather, and other relevant details (see [Listing 2](#) for an enriched XML example).

Data extraction from strings was crucial for exploiting the semantic richness of the description. This approach enhances the ability to perform semantic queries and reveals the numerous latent contexts initially conveyed in an unstructured way ([Giagnolini et al., 2024](#)). Although conversion to RDF could have occurred without this process, it would have resulted in a dataset that could not benefit from its full communicative and semantic potential. Consequently, we would be unable to explore the full representational potential of the models thoroughly.

Regarding the ArchOnto representation, this process had been conducted within the broader EPISA project, which involved extracting information directly from data already migrated into OWL format rather than from the original EAD XML file. This extraction was performed on information derived from various finding aids and different ISAD(G) elements using Natural Language Processing (NLP) techniques ([Varagnolo et al., 2023](#)). Specifically, the process comprised three steps:

- (1) a classifier is run over the national archives information OWL representation;
- (2) for each classified text linked to a document (*E31 Document*), the text and the corresponding document reference code are sent to an information extraction process;
- (3) the information extraction process extracts a set of relations which, using the OWL mapping rules, represent the extracted information in CIDOC CRM linked to the document from which the information was extracted ([Varagnolo et al., 2022](#)).

In the context of the representation in RiC-O, the extraction process involved the combined use of regular expressions ([Python Software Foundation, 2024b](#)) to identify specific scenarios within strings, the lxml library ([Richter, 2024](#)) for manipulating XML tags, and the datetime library ([Python Software Foundation, 2024a](#)) for managing and formatting dates. The extracted information was re-stored in XML format, enriching the initial files with new tags (i.e. <child>, <mother>, <father>, <birth date> and others). Although outside the EAD domain, these tags provided a structured and semantically clear foundation for the conversion process.

This extraction technique was chosen because the strings to be analysed all had the same easily distinguishable structure. Using regular expressions allowed us to quickly achieve reliable results, enabling us to focus directly on the representation requirements, our primary goal. However, this was simply due to the highly homogeneous characteristics of the strings; for larger-scale projects with a wider variety of string types, structures, and text styles to analyse, the application of NLP techniques or artificial intelligence will undoubtedly be necessary ([Giagnolini et al., 2024](#); [Babaei Giglou et al., 2023](#)).

Listing 1. Baptism record of Maria da Silva from the XML EAD file provided by the District Archive of Porto

```
<c level="item" id="000001">
<did>
<langmaterial>Por (português)</langmaterial>
<origination label="RecipientAddress">Lugar de Vilarinha, Freguesia de
Aldoar</ origination>
<physdesc>
<dimensions>120x210mm ; papel</dimensions>
</physdesc>
<physloc>E/20/6/3 - 9.2 - fl. 1, ass. 1</physloc>
<repository>Arquivo Distrital do Porto</repository>
```

```
<unitdate label="UnitDates" type="inclusive" certainty="True/True"
normal="
1707-07-20/1707-07-20">1707-07-20/1707-07-20</unitdate>
<unitid countrycode="PT" repositorycode="PT-ADPRT">000001</unitid>
<unittitle type="Formal">Registo de batismo de Maria </unittitle>
</did>
<scopecontent>
<p>Pais: Manuel Antonio e Ana da Silva Data de nascimento: não
mencionado</p>
</scopecontent>
<odd label="OriginalNumbering">
<p>M2</p>
</odd>
</c>
```

Listing 2. Baptism record of Maria da Silva provided with additional tags derived from the information extraction process

```
<c level="item" id="000001">
<did>
<langmaterial>Por (português)</langmaterial>
<origination label="RecipientAddress">Lugar de Vilarinha, Freguesia de
Aldoar</ origination>
<physdesc>
<dimensions>120x210mm ; papel</dimensions>
</physdesc>
<physloc>E/20/6/3 - 9.2 - fl. 1, ass. 1</physloc>
<repository>Arquivo Distrital do Porto</repository>
<unitdate label="UnitDates" type="inclusive" certainty="True/True"
normal="
1707-07-20/1707-07-20">1707-07-20/1707-07-20</unitdate>
<unitid countrycode="PT" repositorycode="PT-ADPRT">000001</unitid>
<unittitle type="Formal">Registo de batismo de Maria </unittitle>
</did>
<scopecontent>
<p>Pais: Manuel Antonio e Ana da Silva Data de nascimento: não
mencionado</p>
</scopecontent>
<odd label="OriginalNumbering">
<p>M2</p>
</odd>
<child>Maria da Silva</child>
<birth_date> não mencionado</birth_date>
<father>Manuel Antonio</father>
<mother>Ana da Silva</mother>
</c>
```

2.3 XML to RDF conversion

In the third stage, for the conversion of formats, we selected the open-source tool provided by the National Archives of France, namely the RiC-O converter (Francart, 2024).

RiC-O converter is a tool to convert EAD and EAC XML files to RDF datasets conforming to RiC-O, written mostly using XSLT stylesheets and wrapped in a Java command-line application (Clavaud *et al.*, 2023).

Following the guidelines provided in the documentation, we tailored the converter to ensure that it accurately reflected the complexities and nuances of our dataset and to obtain a final RDF file conforming to RiC-O. The main changes to the converter were made to the `ead2rico.xslt`, which is included within the application folder dedicated to the conversion of EAD files and contains all the conversion logic.

The changes mainly involved: inserting the mapping of tags not initially considered by the converter but present in the original EAD file, such as `<phylsoc>` and `<bioghist>`; inserting the mapping of tags that resulted from the data extraction procedure, such as `<father>` or `<mother>`; and creating and modifying mapping scenarios to meet the representation logic needed for our case study. A good example is the handling of dates, which exemplifies the latter case. In the RiC-O converter, dates were by default translated through the adoption of the data properties `<rico:date>`, `<rico:beginningDate>` and `<rico:endDate>`. To match our requirements, dates were instead translated into instances of the class `<rico:Date>`, semanticised through the adoption of object properties such as `<rico:isBirthDateOf>`, `<rico:isCreationDateOf>`, `<rico:isBeginningDateOf>` and `<rico:isEndDateOf>`. All specific references to the National Archives of France, which were present mainly for managing information about the authorities, URIs, and vocabularies, have been removed [2].

3. Results and discussion

The resources generated through the process described in the “Case Study and Methods” Section include:

- (1) Semantic maps based on RiC-O and ArchOnto representing the four identified scenarios (birth event, baptism activity, record description and provenance documentation).
- (2) The Python script for data extraction.
- (3) The edited XML file for format conversion, inclusive of the new extracted information.
- (4) The edited RiC-O converter XSLT file.
- (5) The RDF file containing the dataset compliant with RiC-O.

These materials, together with the original XML EAD file provided by the District Archive of Porto, are available in the official repository of INESC TEC ([Giagnolini and Koch, 2024](#)) to provide a comprehensive view of the methodologies and tools employed, facilitating transparency and reproducibility for further research.

To meet our goal of comparing ArchOnto and RiC-O approaches, we will discuss the characteristics of the semantic maps depicting the four scenarios with both models. Each scenario will be developed and discussed through the representation of data through figures (visible in the section addressing the scenario under consideration) and RDF code snippets, available in [Appendix](#).

3.1 Birth event

[Figure 2](#) illustrates the classes and properties required for the representation of an instance of a birth event in ArchOnto. As for RiC-O, we selected two main possibilities, presented in [Figures 3 and 4](#). The representations can also be seen in the form of RDF snippets, available in [Table A1 in Appendix](#), where two snippets are depicted side by side to facilitate comparison.

The first aspect to analyse is the representation of dates. In ArchOnto (and thus in CIDOC CRM [3]), the way of representing the date is only one and not very straightforward. As we can see in [Figure 2](#), the date must be represented through an instance of the *E52 Time-Span* class, typified by an *ARE6 Date Type* and identified by an *E41 Appellation*, which in turn must be

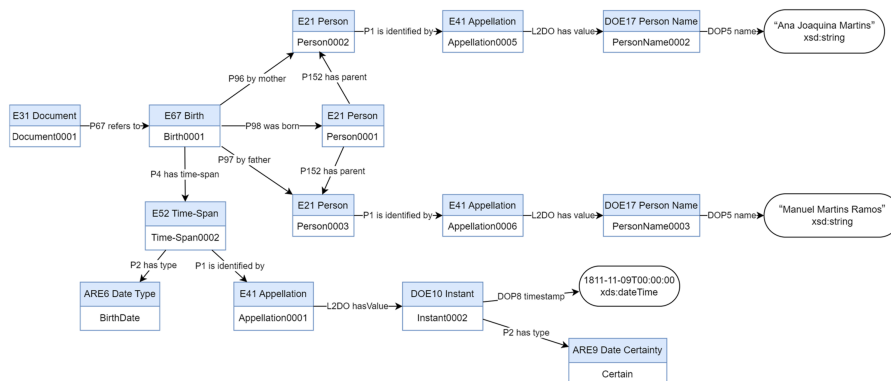


Figure 2. Representation of birth using ArchOnto. Source: Authors' own work

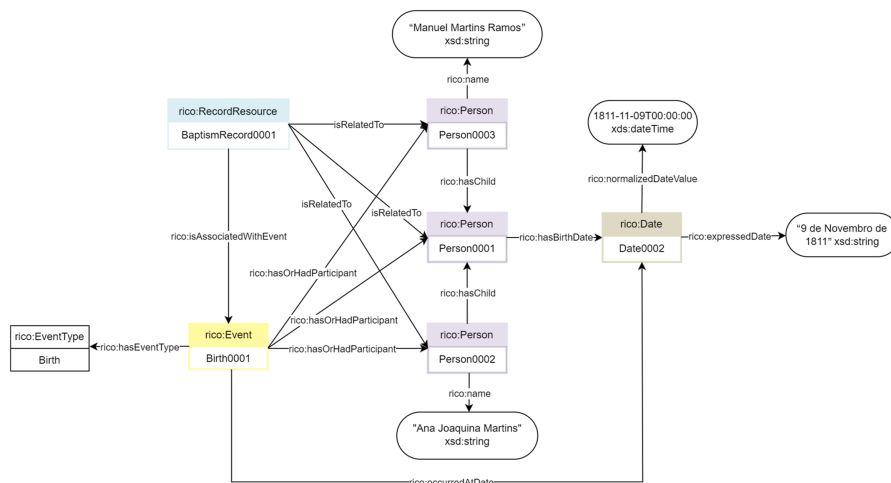


Figure 3. First option to represent a birth event with RiC-O. Source: Authors' own work

associated with an *DOE10 Instant* value that allows the normalised date and its degree of certainty to be determined.

RiC-O allows considerable flexibility in the expression of dates to be associated with a resource, ranging from the adoption of simple data properties such as `<rico:date>` to the expression through the use of the `<rico:Date>` class. In the latter case, as is evident from the drawings in Figure 3 [4] and Figure 4, one can typify the date through the object property that connects it to the resource, such as `<rico:hasBirthDate>`. Additionally, it is possible to specify further characteristics through data properties associated with the `<rico:Date>` class itself, such as `<rico:dateQualifier>` and `<rico:normalizedDateValue>`.

Compared to ArchOnto, RiC-O can represent the dates more intuitively. The primary way to define the type of date is through the use of semantically explicit object properties, such as `<rico:hasBirthDate>`,

`<rico:hasCreationDate>`, `<rico:hasDateOfOccurrence>`, and `<rico:hasEndingDate>`. Note that the `<rico:DateType>` class also exists in RiC-O. However, the scope note makes explicit

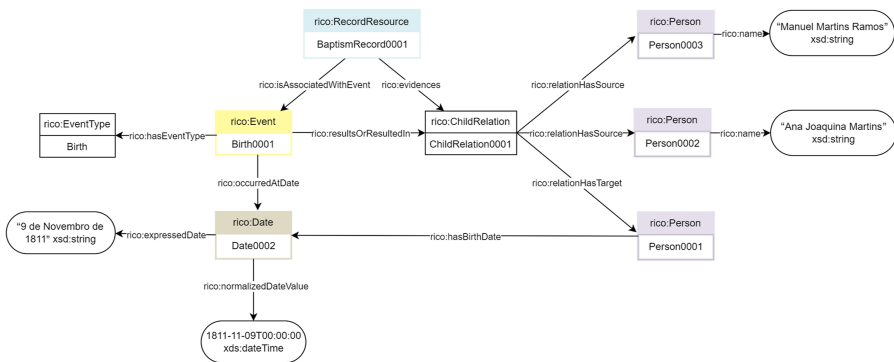


Figure 4. Second option to represent a birth event with RiC-O. Source: Authors' own work

that this class “may be used to categorise a Date as a single date, a range of dates, or a set of dates or subcategories of these broad types” and that it “should not be confused with date relationships defined to link a Date entity and any other entity (such as RiC-R069 “is the start date of”)” (International Council on Archives Expert Group on Archival Description, 2023a, p. 48).

Consequently, although the range of object properties is quite broad, the impossibility of using `<rico:DateType>` to typify the date poses a challenge. Defining the date type for phenomena slightly outside traditional archival contexts may result in either an inadequate representation of the relation’s semantics (e.g. opting for the most generic `<rico:isDateAssociatedWith>`) or defining a non-canonical approach such as using the generic class `<rico:Type>` or the data property `<rico:type>`.

Another interesting condition to analyse is the representation of the relationship between parents and child [5]. RiC-O allows representing it in two ways: either very directly, through the object property `<rico:hasChild>` (and its inverse `<rico:isChildOf>`), which connects the instances of `<rico:Person>` representing the parents and the child, or through the addition of the class `<rico:ChildRelation>`, which associates the parents as the source and the child as the target of the relationship. Nevertheless, considering that adding the `<rico:ChildRelation>` complicates the representation without bringing any particular benefit to its semantics, the data were modeled according to the schema defined in Figure 3. Additionally, one should note that RiC-O has no direct way to state the role of “mother” or “father” since they are more generically represented as “parents”. However, if necessary, one could be more specific by identifying their `<rico:DemographicGroup>` and specify the gender or biological sex of each person.

The semantic scope of the relation expressed in these two ways would be sufficient to witness the child’s birth and connect it to the record resource that mentions it. However, it is possible to further specify this contingency by identifying the birth as an actual event using the `<rico:Event>` class, connected directly to the instances of the people (Figure 3) or the `<rico:ChildRelation>` (Figure 4). In any case, the child is directly associated with a date of birth via the object property `<rico:hasBirthDate>`; the same date, of course, will be the one on which the birth event occurred.

In ArchOnto, the representation of the birth event appears as a mix of RiC-O’s ones (Figure 2). *E67 Birth* is a specific class used to attest the birth event and is linked to instances of people through explicit object properties: *P96 by mother* and *P97 by father* to link it to the parents, and *P89 was born* for the child. Between them, the child and parents can be associated through the property *P152 has parent*. The main difference is that dates in ArchOnto cannot be directly related to instances of people, as in RiC-O. Still, it must necessarily be associated

through an intermediate node, making the *E67 Birth* presence essential to describe the birth of someone comprehensively.

Although we have not represented it in the data or the drawings, one could also do an analysis to identify the child's siblings. However, while with RiC-O, it is possible to relate siblings either directly, through the object property <rico:hasSibling>, or through the class <rico:SiblingRelation>, in ArchOnto, there is no way to define such a relation. The only way to trace back the information is to identify all the *E67 Birth* instances linked to both parents.

3.2 Baptism activity

The baptism activity is the central focus of this archival record. To represent this activity, it is necessary to consider the people who participated, their roles, and the date it happened. Figure 5 illustrates the classes and properties required for the representation of an instance of a baptism in ArchOnto. As for RiC-O, we selected two main possibilities, presented in Figures 6 and 7. The representations can also be seen in the form of RDF snippets, available in Table A2 in Appendix, where two snippets are depicted side by side to facilitate comparison. In ArchOnto, as can be seen in Figure 5, the baptism activity is represented through the *E7 Activity* class, which has a type specified through *ARE16 Event Type*. This activity involves several actors defined through *E21 Person*. To represent these actors, it is necessary to consider their role in the present activity through an N-ary relation. This N-ary representation indicates that the person (*E21 Person*) has a specific role (*ARE8 Role*) in the baptism activity (*E7 Activity*), i.e. the role of godfather or godmother. Anyway, if the role is not known, one can generically state that the person (*E21 Person*) participated in the activity (*E7 Activity*) by using the object property *P11 had participant*.

RiC-O also proves to be quite flexible in describing the birth activity. The representation shown in Figure 6 includes the use of <rico:PerformanceRelation> class, which allows linking the <rico:Person>, i.e. the godfather, to the role he plays in the activity, which is identified as the context of the relationship. This representation is not intuitive as no object property explicitly signifies "has role" in RiC-O.

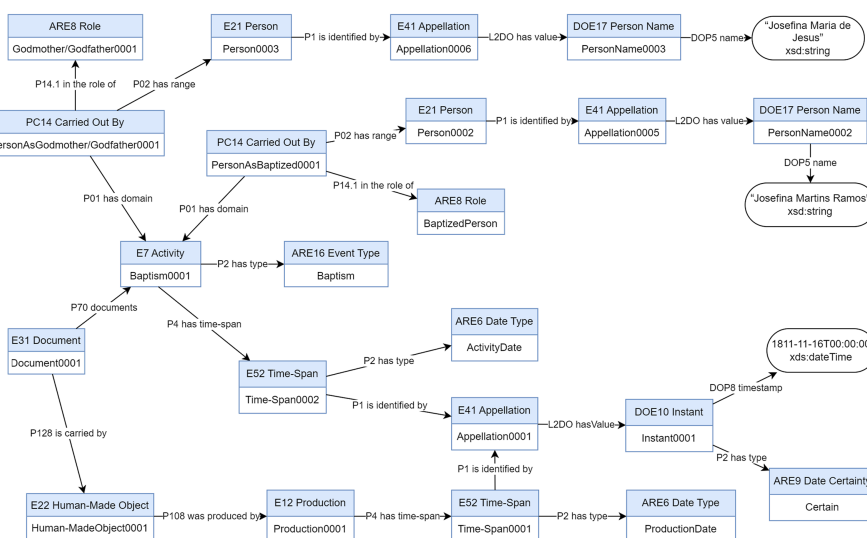


Figure 5. Representation of the baptism activity with ArchOnto. Source: Authors' own work

directly to the person class using the data property <rico:name> is possible. As there is no need to specify further details concerning the agents' names, we have decided to represent the data in RiC-O using this second option.

Additionally, as seen from both representations, we should add that the record date that documents the baptism activity corresponds to the date of the baptism itself (since the recording typically co-occurred).

3.3 Record resource description

Figure 8 illustrates the classes and properties required for the representation of an instance of an archival record in ArchOnto. As for RiC-O, the result is presented in Figure 9. The two representations can also be seen in the form of RDF snippets, available in Table A3 in Appendix, where the two snippets are depicted side by side to facilitate comparison. The ArchOnto's representation of the archival resource is developed through the use of three classes: *E31 Document*, *E33 Linguistic Object*, and *E22 Human-Made Object*, as can be seen in Figure 8. The first, *E31 Document*, is the central node of the representation and connects the linguistic-textual

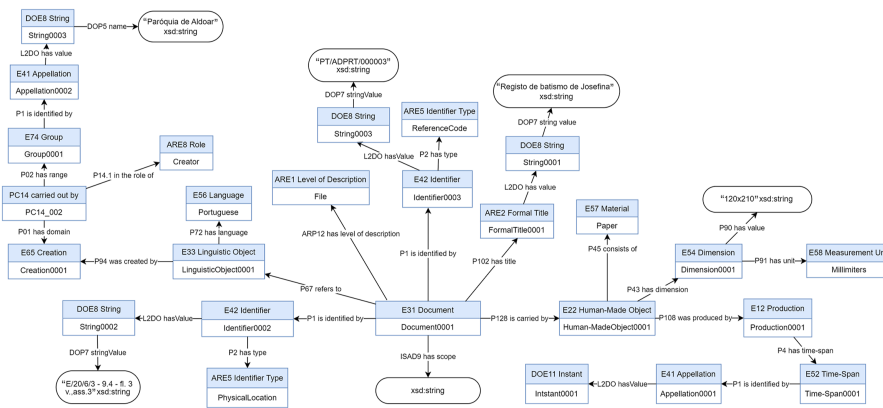


Figure 8. Representation of the record resource and its context in ArchOnto. Source: Authors' own work

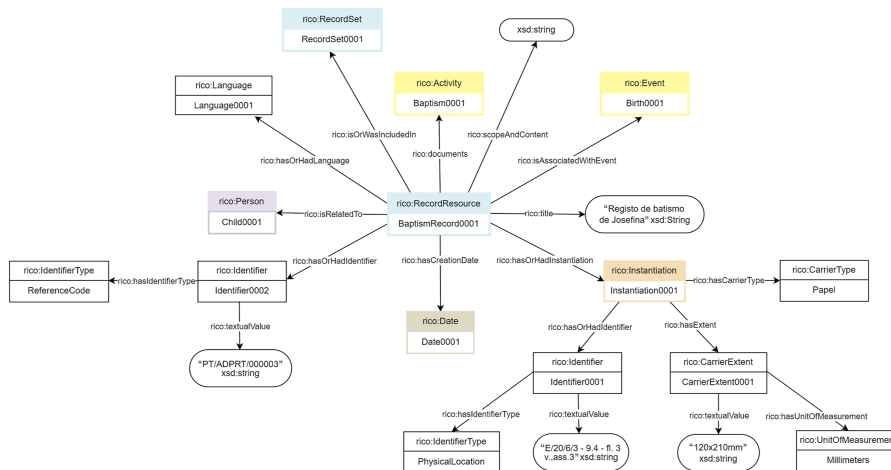


Figure 9. Possible representation of record resource and its context in RiC-O. Source: Authors' own work

content of the document (*E33 Linguistic Object*) to its realisation in a physical object (*E22 Human-Made Object*). Linked to the *E31 Document* are the main identification coordinates of the archival document, i.e. the identifiers, the level of description and the formal title.

RiC-O's representation of the archival resource, as shown in [Figure 9](#), is twofold in the setting of the `<rico:RecordResource>` – i.e. the informative content – and its physical materialisation(s), the `<rico:Instantiation>`. Accordingly, all document characteristics related to its informative content – such as its language or the activity it documents – are linked to the record resource, while all the strictly physical characteristics – such as the physical location and the carrier extent – are connected to its instantiation.

The representation of baptism records exemplifies one of the primary reasons for the introduction of the `<rico:RecordResource>` class as a superclass of `<rico:Record>` and `<rico:RecordPart>`, i.e. the potential challenge in distinctly categorising an archival object as either a record or a record part ([Clavaud and Wildi, 2021](#), p. 8). Indeed, while baptism records have independent and autonomous informative content, making them eligible to be represented as `<rico:Record>`, they are physically part of a book that seamlessly collects them, making them also potentially representable as `<rico:RecordPart>`. For this reason, avoiding making sharp and debatable choices on one side or the other, we decided to represent the individual baptism records as `<rico:RecordResource>`. The baptism record is linked to class instances resulting from the information extraction performed on the ISAD(G)'s "Scope and Content" element, such as the newborn child or the birth event. However, for clarity and accessibility, it was also preferred to keep the "Scope and Content" element in its integrity, preserving it as a string in the data property `<rico:scopeAndContent>`.

The first element to consider when comparing the two representations is the use of classes to identify the archival document (as an object and information). With the use of three classes – *E22 HumanMade Object*, *E33 Linguistic Object*, and *E31 Document* – ArchOnto presents a more intricate structure. The semantic difference between the last two classes, in particular, is rather subtle: while both refer to informative content, a *E31 Document* is an abstract information object that may exist in various formats (text, image, video), whereas a *E33 Linguistic Object* specifically pertains to content expressed through language.

This meticulous categorisation captures the detailed nature and relationships of records, but this complexity can be burdensome, requiring significant efforts to understand and implement. The simpler bipartition between `<rico:RecordResource>` and `<rico:Instantiation>` in RiC-O may facilitate a more intuitive understanding and management of records just by distinctly separating the intellectual content from its physical carrier.

Considering that this modelling is a novelty compared to previous standards ([Clavaud and Wildi, 2021](#), p. 8), and archivists and researchers still have to get used to this approach of representing information, this simplicity may be crucial. A more straightforward framework means quicker and more effective understanding and implementation.

CIDOC CRM would theoretically allow for a direct link between a *E33 Linguistic Object* and a *E22 Human-Made Object*, making its representation more comparable and aligned with RiC-O's. In this scenario, information associated with the *E31 Document* must be distributed across these two classes. Nevertheless, even if this approach is feasible, it would exclude archival documents that do not strictly qualify as *E33 Linguistic Object* (i.e. photographs, drawings, video and audio recordings, etc.). On the other hand, omitting the *Linguistic Object* class by keeping only the *E31 Document* would prevent the indication of the document's language, which in CIDOC CRM can only be associated with the *E33 Linguistic Object* class.

Another interesting aspect is the handling of ISAD(G)'s element "Level of Description". This is, in fact, one of the extensions made by the ArchOnto model since CIDOC CRM does not provide for it, being born for the description of museum objects ([Koch et al., 2020](#); [Koch et al., 2023b](#)). ArchOnto meets this requirement by introducing the class ARE1 Level of Description, which is linked to the document via its object property ARP12 has level of description.

RiC-O expresses the concept differently, first of all identifying the tripartition of the `<rico:RecordResource>` into the three subclasses `<rico:Record>`, `<rico:RecordPart>`, and `<rico:`

RecordSet>. Referring to the levels of description proposed by ISAD(G), the record is generally understood as an “item”. All levels that, theoretically or in fact, group records together are record sets. Each record set, if necessary, can be further typified by identifying a <rico:RecordsetType>.

Delving further into detail, we note that ArchOnto introduces an aspect not considered by CIDOC CRM nor by RiC-O : a specification of the title that could be original or attributed. ArchOnto indeed extended the CIDOC CRM class E35 Title with the two subclasses ARE2 Formal Title and ARE3 Supplied Title. In RiC-O, there are no such specificities expressed in classes, and the only way to define this characteristic would seem to be the association of the class <rico:Title> with the data property <rico:type> [9]. This may seem almost trivial, but it is not if we consider that many archival resources do not have an original denomination but are titled by the archivist.

In a context where the representation of informational content of the document is separate from the one of its physical form, it becomes increasingly essential to specify the title’s origin, as it may also differ between these two representations.

3.4 Provenance documentation

In the LOD context, the attestation of data provenance is crucial for ensuring reliability and traceability (Tomasi, 2017).

At a fundamental level, an ontology should be sufficiently expressive to capture essential provenance information within a simple triple, ensuring that basic informational needs are met even in the most straightforward representations, i.e. to define who is the author of a specific resource. Nevertheless, more than relying on a simple triple is often required. While traditional RDF triples are adequate for representing basic facts, they lack the expressive power to capture the complex contextual information needed to fully attest to the provenance of data (Sikos and Philp, 2020).

Provenance often involves not just the data itself but also the circumstances under which it was generated and modified, including details such as the agents involved, the time, place, and method of data collection, and the specific source or evidence supporting the data.

In our case study, the extraction of entities from the text fields, the attribution of relational semantics between them, and the format conversion effectively constitute an interpretative act of the textual contents (Giagnoloni *et al.*, 2024). Therefore, the new finding aid expressed in linked data must be explicitly identified as the result of a new interpretative act, distinct from the archival description activity that produced the original EAD finding aid. So, the files resulting from the conversion should ideally be accompanied by a series of additional triples that explicitly declare their provenance by attesting their origin, their production methods, and, ultimately, the attribution of responsibility (Tomasi, 2023).

To address this issue, adopting RDF reification, n-ary relations, rdf-star, or named graphs are the most used strategies (Sikos and Philp, 2020; Rupp *et al.*, 2022).

ArchOnto has not yet defined a specific approach regarding the representation of provenance. However, since ArchOnto is built on the CIDOC CRM ontology, we explored other models compatible with CIDOC CRM to identify existing solutions for representing this type of information. Our investigation revealed that CRMdig (FORTH and CRM SIG, 2016) already offers a framework for provenance representation. Accordingly, a study was undertaken to evaluate the feasibility of integrating CRMdig into ArchOnto.

CRMdig – currently in version 4.0 – is an ontology and RDF Schema to encode metadata about the steps and methods of production (“provenance”) of digitisation products and synthetic digital representations such as 2D, 3D or even animated models created by various technologies (FORTH and CRM SIG, 2016) [6].

On the other hand, EGAD in the official RiC-CM documentation emphasises the importance of documenting the description by identifying at least four layers of context: documenting the holding agent, establishing the position responsible for processing and

describing records, documenting the archival description record itself, and documenting the evidence for assertions made in the description record ([International Council on Archives Expert Group on Archival Description, 2023a](#)).

RiC-O consequently makes it possible to make these four contexts explicit. In particular, provenance management is facilitated through n-ary relations, which provide a high degree of expressiveness. This approach stems from the decision to reify relations within the <rico:Relation> class and its extensive set of subclasses. These subclasses, such as <rico:AuthorshipRelation>, <rico:AccumulationRelation>, or <rico:MigrationRelation>, are designed to map most of the relationships that arise in documentary and curatorial contexts. However, the need to find solutions to record the provenance of more complex assertions has already emerged in the community. The issue has been set as a milestone for future RiC-O developments ([ICA-EGAD, 2024](#)).

Given these premises, in [Figures 10 and 11](#), it is possible to observe two possible representations of data provenance through ArchOnto and RiC-O, which can also be seen in the form of RDF snippets, available in [Table A4 in Appendix](#), where the two snippets are depicted side by side to facilitate comparison.

The provenance representation in ArchOnto, using CRMdig, is mainly used to indicate the data extraction and migration event from a finding aid in EAD to an RDF one [7]. It was necessary to start with the event representation, defined through the *D7 Digital Machine Event* class, which has the event type represented by the *ARE16 Event Type* class of ArchOnto. The *D7 Digital Machine Event*, from CRMdig, “comprises events on physical-digital devices following a human activity that intentionally caused its immediate or delayed initiation and resulted in creating a new instance of the *D1 Digital Object* on behalf of the human actor” ([FORTH and CRM SIG, 2016](#)).

Using the event as a basis, it was possible to indicate, through an n-ary relationship, who was the archivist responsible for the conversion using the RiC-O Converter. To represent the n-ary relationship, creating a class that links all the elements in the relation is necessary – i.e. *PC14 Carried Out By*. With this class created, we can indicate that a person (*E21 Person*) was an archivist (*ARE8 Role*) in the event (*D7 Digital Machine Event*). In turn, to represent the use of the RiC-O Converter, it was necessary to use the *L23 used software or firmware* property, which links the *D7 Digital Machine Event* to a *D14 Software*.

As we intend to represent the transformation of an archival record in EAD to an RDF document, it is essential to show this change, which is done through the properties *L10 had input* and *L11 had output*, which links each one to an instance of *D1 Digital Object*. Each *D1*

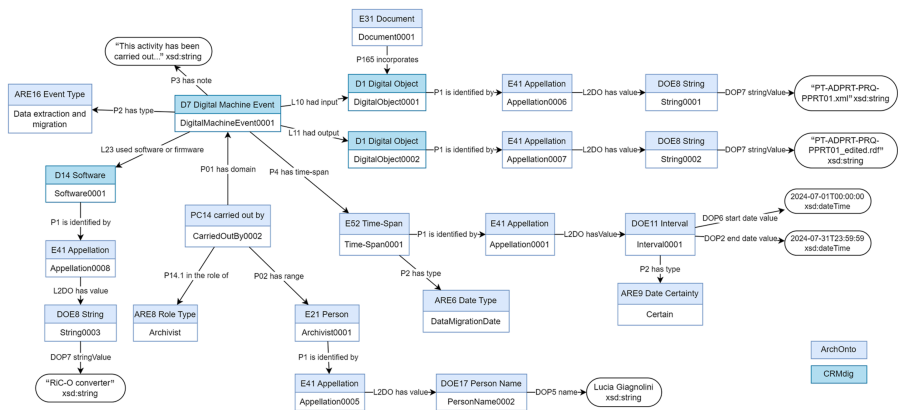


Figure 10. Representation of data extraction and conversion in ArchOnto. Source: Authors’ own work

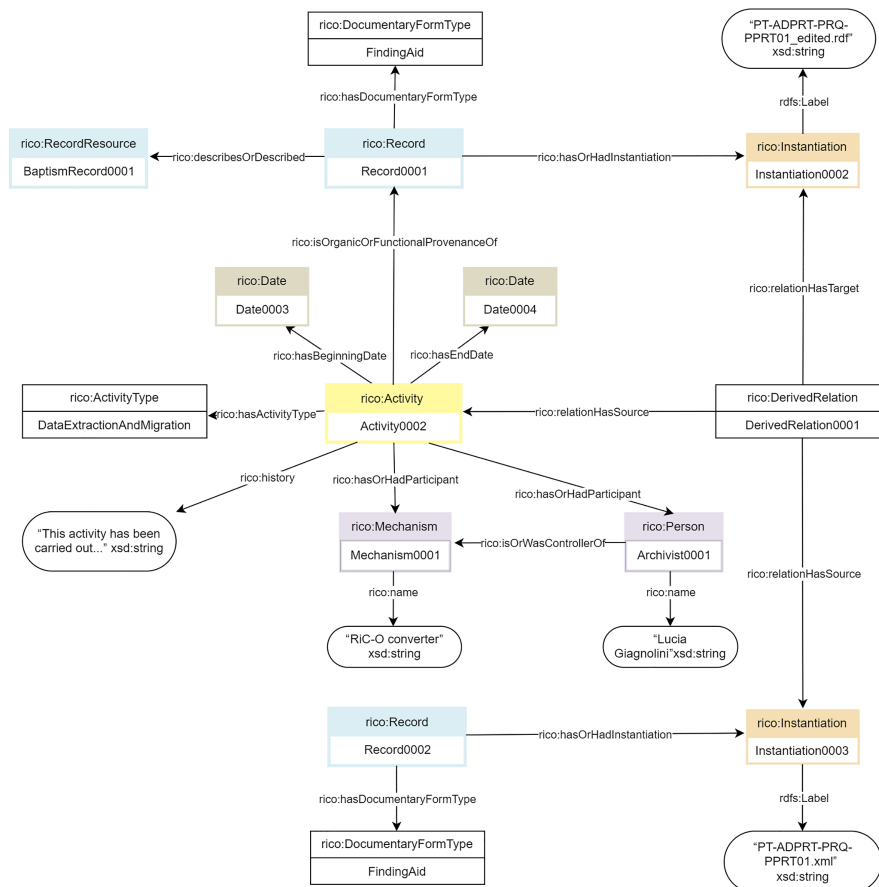


Figure 11. Representation of data extraction and conversion in RiC-O. Source: Authors' own work

Digital Object represents the EAD file as the input and the RiC-O as the output. To fully describe the event, ArchOnto's data property *P3 has note* was used.

The application of CRMdig within ArchOnto offers several advantages, particularly in its ability to represent the conversion of data formats directly and coherently. The model's consideration of the digital environment as the context of such events aligns well with contemporary archival practices, where digital transformation is increasingly prevalent. Moreover, the ability to specify a *D7 Digital Machine Event* as being carried out "on behalf of the human agent" is a significant strength. This aspect effectively captures the intentionality behind the archivist's actions and the interaction dynamics with the software. However, a notable drawback is that CRMdig is not entirely stable yet, which could lead to potential issues in long-term implementation and interoperability. Despite this, integrating CRMdig into ArchOnto proved to be a milestone for future development.

As for the representation in RiC-O, following RiC-CM specifications, finding aids can be represented as `<rco:Record>` ([International Council on Archives Expert Group on Archival Description, 2023a](#)). Consequently, in the drawing, the EAD finding aid and the RDF one are

represented as two `<rico:Record>` having `<rico:DocumentaryFormType>` set as “FindingAid”. To document the relationship between them, the class `<rico:DerivationRelation>` was identified, which connects the instantiation of one record to a second instantiation from which it is derived. To further put this into context, the data extraction and migration activity is identified as the organic provenance of the record representing the description in RDF and the context of the `<rico:DerivationRelation>`. The activity – which has a start date and an end date – has two participants, the archivist and the software she uses, the RiC-O converter; the former is in the role of “controller” of the latter. The data property `<rico:history>` was also given value to describe the activity in detail.

Although this representation achieves its objective, some areas of shadow and redundancy are worth discussing. According to the RiC-O specifications, the `<rico:DerivationRelation>` can only connect instances, lengthening and complicating our representation. However, it is evident that informative content can also be derived from other informative content, thus allowing one record to be derived from another. Expanding the scope of `<rico:DerivationRelation>` to include records could explore ways to simplify the model without sacrificing the richness of information, making it more usable in practical applications and less prone to redundancy.

Additionally, the relation between the archivist and the software used to perform the activity, represented by the object property `<rico:isOrWasControllerOf>`, does not accurately capture the dynamics of the relationship. While it is certainly valid to identify the software as an agent playing a role in the activity, it would be more appropriate to introduce a property that precisely defines the archivist’s use of the software, such as depicted in CRMdig. This enhancement would provide a clearer understanding of the interaction, emphasising the role of the software as a tool the archivist utilises in executing archival tasks.

One final observation concerns the definition of the activity of data extraction and conversion as the context for both the derivation relationship between finding aids and, consequently, the creation of the finding aid in RDF. While the property `<rico:HasOrganicOrFunctionalProvenance>` correctly captures the fact that the finding aid results from a specific activity, in the sense that the document does not merely document the activity but is the outcome of it, the object property `<rico:relationHasContext>` is less meaningful in this context. In fact, `<rico:relationHasContext>` is semantically weak because the activity of data extraction and migration is what determines the relationship itself, rather than being a secondary factor, as implied by the RiC-O definition. Using `<rico:relationHasSource>` would make more sense, but in this case, the relationship is meant to connect the document from which the finding aid is derived. Alternatively, one could consider using `<rico:HasOrganicOrFunctionalProvenance>` again, but unfortunately, this property does not have `<rico:Relation>` as its domain.

The two diagrams provide an initial approach to address the challenge of tracing the provenance of format conversion activities, but both need further developments [8]. The need for more robust and precise models becomes evident as format conversion becomes an increasingly common practice within archival institutions particularly for migrating XML inventories to RDF (a trend anticipated to grow in the coming years).

A significant positive aspect is that, for both CIDOC CRM and RiC-O, the respective communities have recognised the necessity of representing more complex situations, and dedicated working groups actively enhance these models. In addition to considering the adoption of more complex strategies, such as the aforementioned named graphs or rdf-star, one promising direction could involve the integration or alignment with ontologies such as PROV-O (Belhajjame *et al.*, 2013) or HiCO (Daquino, 2020; Tomasi, 2017), created explicitly to deal with contexts of origin and applied quite extensively. This would enable the use of more specific terminologies and facilitate ontological alignment and interoperability across these models regarding provenance. Such efforts could lead to more comprehensive and refined frameworks that meet the evolving demands of data provenance in archival practices.

3.5 General discussion

The analysis so far has allowed us to identify how the two models represent archival records, relations and contexts highlighting their characteristics in terms of flexibility, complexity, general use, provenance expression, users and developers communities. Each scenario identified by this study highlighted specific differences and similarities in representation and initiated general considerations summarized in [Table 1](#).

Overall, our results are in line with those of [Oliveira et al. \(2024\)](#), who compared CIDOC CRM and RiC-O on the basis of the definitions of classes and properties. Indeed, ArchOnto benefits from a well-established and widely used core model (CIDOC CRM) that integrates effectively with other GLAM domains. The wide range of its applications prompted ISO to adopt CIDOC CRM as a standard, subsequently adapting the model's concepts for ISO 21127:2014. However, the level of abstraction required to coordinate multiple domains comes at the cost of reduced specialisation for archival needs, which ArchOnto extensions aim to supply. In contrast, RiC-O is not yet integrated with other GLAM domains and remains a relatively recent development, with a model that is still stabilising but that is specifically tailored for archives. The level of abstraction of CIDOC CRM makes it hardly flexible and more cumbersome compared to RiC-O that clearly offers a more versatile and streamlined representation. In this respect, having set the comparison on a specific case study, in addition to the study of [Oliveira et al. \(2024\)](#), allowed us to explore the property chains needed in CIDOC CRM to express concepts that with RiC-O only require a triple, such as the expression of a birthdate. This is also evident at quick glance by looking at [Tables A1, A2, A3 and A4 in Appendix](#), where code snippets for the two representations are shown side by side. Both frameworks share some limitations, as the expression of provenance, yet they are both maintained by active communities of users and developers, indicating ongoing support and improvements.

4. Conclusions and future work

The RiC-O model demonstrates a clear advantage in archival contexts, where its straightforward applicability makes it particularly well-suited for managing and representing archival records. Its flexibility is one of its most valuable assets, allowing it to adapt to different archival institutions' varied and specific needs. This adaptability ensures that RiC-O can be employed across various archival scenarios, from simple cataloguing tasks to complex data management processes. However, this flexibility also introduces a drawback: it can lead to inconsistencies and fragmentation when different institutions implement the model in ways that diverge significantly in the representation of the same concepts. To truly overcome the pervasive issue of data silos, there is a growing need for convergence around a unified model and representation patterns that can serve as a standard across the archival community.

Table 1. Comparison of ArchOnto and RiC-O: Pros and Cons

| ArchOnto | RiC-O |
|--|---|
| Less flexible representation | More flexible representation |
| More cumbersome representation | Less cumbersome representation |
| More integrated with other GLAM domains | Less integrated with other GLAM domains |
| The core base (CIDOC CRM) is developed specifically for the museums domain | Developed specifically for the archives domain |
| The core base (CIDOC CRM) is a well-established model | Recent model and ontology |
| Lack of completeness for the expression of provenance | Lack of completeness for the expression of provenance |
| Active users and developers community | Active users and developers community |
| Source(s): Authors' own work | |

In contrast, CIDOC CRM offers a robust framework that encompasses a wider and more diverse range of domains beyond archives. This model provides the tools to represent complex relationships and entities across various cultural heritage contexts, making it a powerful instrument for interdisciplinary projects. However, its complexity can pose challenges, particularly for institutions needing more resources or expertise to implement and leverage its capabilities thoroughly. Despite these challenges, CIDOC CRM remains the cornerstone for cross-domain interoperability, particularly given its successful integration with the library domain through LRMoo.

This article also aims to emphasise the importance of converting finding aids from XML EAD to linked data models to improve the Findability, Accessibility, Interoperability, and Reusability (FAIRness) of archival data. In this sense, we would also like to highlight the significance of performing information extraction operations rather than merely executing a direct conversion. The objective is to explicitly capture as much contextual information as possible from textual fields. By extracting and structuring relevant entities and their connections, the conversion process can foster a more nuanced, semantically rich representation that aligns with the interconnected nature of linked data. In doing so, archival institutions can fully exploit the advantages of linked data technologies for enhanced discovery, interoperability, and reusability across domains.

Future research and development should focus on aligning RiC-O and CIDOC CRM within the framework of LOD and FAIR principles. Such alignment is not only desirable but advantageous, as it aligns with the long-term goals of the EGAD (Clavaud, 2023), which aims to create more integrated and interoperable archival descriptions. By pursuing this alignment, the archival and broader cultural heritage communities can work towards a more cohesive and effective approach to data and its provenance, bridging different domains while maintaining the specificities of each.

Notes

1. It is worth noting that the OWL files reflect data represented with an outdated version of ArchOnto, so the representation is not fully compliant with the semantic maps, which were drawn to maximise the representational reach of the model in its latest version. The files are available in Pires *et al.* (2023).
2. Note that the resource URIs have been included for illustrative purposes. In implementing resources within a LOD network, the URIs should be constructed accordingly, following best practices such as those outlined by the W3C Working Group <https://www.w3.org/TR/ld-bp/>
3. From here on, any considerations made on ArchOnto are also valid for CIDOC CRM. When not, differences are highlighted.
4. Note that for the drawings representing RiC-O classes and properties, the colours of the palette identified in the official RiC-CM documentation were used.
5. Since we do not have explicit information about the relation between the parents, we cannot assume that they are married, so the relation between them is not depicted.
6. As the latest version is still unstable, version 3.2.1 was considered in this work.
7. Note that the representation in ArchOnto with CRMdig focuses exclusively on the extraction and conversion operations carried out within the scope of this article, and not those previously executed for the EPISA project. This decision was made deliberately, as we were not the authors of the previous operations, and attempting to represent them could introduce the risk of inaccuracies.
8. Please note that the absence of this representation in the ArchOnto OWL files can be attributed to the fact that this aspect was not addressed in prior research efforts. In contrast, the RiC-O files lack this representation because the existing approach still needs to achieve an entirely satisfactory level of completeness.
9. In the current latest version of RiC-O (v. 1.1), released after the submission of this article, the class `<rico:TitleType>` has been actually added.

10. The work presented in this paper refers to RiC-O version 1.0 and ArchOnto version 0.9, which were the latest versions at the time of submission of the paper.

References

- Amini, R., Norouzi, S.S., Hitzler, P. and Amini, R. (2025), "Towards complex ontology alignment using large language models", *Knowledge Graphs and Semantic Web: 6th International Conference, KGSWC 2024*, Paris, France, December 11-13, 2024, Springer-Verlag, Berlin, Heidelberg, pp. 17-31, doi: [10.1007/978-3-031-81221-72](https://doi.org/10.1007/978-3-031-81221-72).
- Arquivo Distrital do Porto (2013), "Arquivo distrital do Porto website", available at: <https://pesquisa.adporto.arquivos.pt/> (accessed 20 December 2024).
- Arquivo Distrital do Porto (2015), "Registos de baptismos – record description", available at: <https://pesquisa.adporto.arquivos.pt/details?id=488458> (accessed 20 December 2024).
- Arquivo Distrital do Porto (2022), "Par'ouquia de aldoar website – record description", available at: <https://pesquisa.adporto.arquivos.pt/details?id=488455> (accessed 20 December 2024).
- Babaei Giglou, H., D'Souza, J. and Auer, S. (2023), "LLMs4OL: large language models for ontology learning", in Payne, T.R., Presutti, V., Qi, G., Poveda-Villalón, M., Stoilos, G., Hollink, L., Kaoudi, Z., Cheng, G. and Li, J. (Eds), *The Semantic Web – ISWC 2023*, Springer Nature Switzerland, pp. 408-427.
- Belhajjame, K., Cheney, J., Corsar, D., Garijo, D., Soiland-Reyes, S., Zednik, S. and Zhao, J. (2013), "PROV-O: the PROV ontology", available at: <https://www.w3.org/TR/prov-o/>
- Clavaud, F. (2023), "Transform into extension of CIDOC CRM", *GitHub comment*, Issue #50, ICA-EGAD/RiC-O, available at: <https://github.com/ICA-EGAD/RiC-O/issues/50#issuecomment-1508932062>
- Clavaud, F. and Wildi, T. (2021), "ICA records in contexts-ontology (RiC-O): a semantic framework for describing archival resources", *Linked Archives 2021: Proceedings of Linked Archives International Workshop 2021 Co-Located with 25th International Conference on Theory and Practice of Digital Libraries (TPDL 2021) CEUR Workshop Proceedings (CEUR-WS.Org)*, Vol. 79, pp. 79-92, available at: <https://enc.hal.science/hal-03965776>
- Clavaud, F., Francart, T. and Charbonnier, P. (2023), "RiC-O converter: a software to convert EAC-CPF and ead 2002 XML files to RDF datasets conforming to records in contexts ontology", *Journal on Computing and Cultural Heritage*, Vol. 16 No. 3, pp. 1-13, doi: [10.1145/3583592](https://doi.org/10.1145/3583592).
- Daquino, M. (2020), "The historical context ontology (HiCO)", available at: <https://marilenadaquino.github.io/hico/> (accessed 20 December 2024).
- Daquino, M. (2021), "Linked open data native cataloguing and archival description", *JLIS.it*, Vol. 12 No. 3, pp. 91-104.
- FORTH and CRM SIG (2016), "Definition of the CRMdig: an Extension of CIDOC-CRM to support provenance metadata", available at: https://www.cidoc-crm.org/crmvig/sites/default/files/CRMdig_v3.2.1.pdf
- Francart, T. (2024), "RiC-O converter", available at: <https://archivesnationalesfr.github.io/rico-converter/en/> (accessed 20 December 2024).
- Giagnolini, L. and Koch, I. (2024), "Semantic representation of the Registos de Baptismos da Par'ouquia de Aldoar", Porto, Portugal [Data set], doi: [10.25747/15YG-GD86](https://doi.org/10.25747/15YG-GD86).
- Giagnolini, L., Bonora, P. and Tomasi, F. (2024), "Affinare il contesto: estrazione di informazioni strutturate per l'arricchimento dei contesti archivistici", in Silvestro, A.D. and Spampinato, D. (Eds), *Me.Te. Digitali. Mediterraneo in rete tra testi e contesti, Proceedings del XIII Convegno Annuale AIUCD2024*, Catania, Quaderni di Umanistica Digitale, AIUCD, available at: <https://amsacta.unibo.it/id/eprint/7927/>
- ICA-EGAD (2024), "Discussion on refining the definition of an attribute for a legal status", *GitHub*, Issue #56, available at: <https://github.com/ICA-EGAD/RiC-O/issues/56> (accessed 14 August 2024).

- ICOM and CIDOC CRM SIG (2024), “Volume A: definition of the CIDOC conceptual reference model version 7.1.3”, available at: https://www.cidoc-crm.org/sites/default/files/cidoc_crm_version_7.1.3.pdf
- IFLA LRMoo Working Group and CIDOC CRM SIG (2024), “LRMoo object-oriented definition and mapping from the IFLA Library Reference Model”, available at: https://www.cidoc-crm.org/frbroo/sites/default/files/LRMoo_V1.0.pdf
- Institute for Systems and Computer Engineering of Porto (2024), “EPISA project website”, available at: <https://episa.inesctec.pt/>
- International Council on Archives Expert Group on Archival Description (2023a), “Records in contexts – conceptual model, international council on archives”, Paris, France. Version 1.0, available at: <https://github.com/ICA-EGAD/RiC-CM/issues>
- International Council on Archives Expert Group on Archival Description (2023b), *Records in Contexts: Foundations of Archival Description (Version 1.0)*, International Council on Archives, available at: <https://github.com/ICA-EGAD/RiC-CM/issues>
- Koch, I., Ribeiro, C. and Teixeira Lopes, C. (2020), “ArchOnto, a CIDOC-CRM-based linked data model for the Portuguese archives”, in Hall, M., Meřcun, T., Risse, T. and Duchateau, F. (Eds), *Digital Libraries for Open Knowledge*, Springer International Publishing, Vol. 12246, pp. 133-146, doi: [10.1007/978-3-030-54956-5_10](https://doi.org/10.1007/978-3-030-54956-5_10).
- Koch, I., Pires, C., Lopes, C.T., Ribeiro, C. and Nunes, S. (2023a), “From ISAD(G) to linked data archival descriptions”, in Alonso, O., Cousijn, H., Silvello, G., Marrero, M., Lopes, C.T. and Marchesin, S. (Eds), *Linking Theory and Practice of Digital Libraries*, Springer Nature Switzerland, Vol. 14241, pp. 303-309, of Lecture Notes in Computer Science.
- Koch, I., Teixeira Lopes, C. and Ribeiro, C. (2023b), “Moving from ISAD(G) to a CIDOC CRM-based linked data model in the Portuguese archives”, *ACM Journal on Computing and Cultural Heritage*, Vol. 16 No. 4, pp. 1-21, doi: [10.1145/3605910](https://doi.org/10.1145/3605910).
- Liu, X., Tong, Q., Liu, X. and Qin, Z. (2021), “Ontology matching: state of the art, future challenges, and thinking based on utilized information”, *IEEE Access*, Vol. 9, pp. 91235-91243, doi: [10.1109/access.2021.3057081](https://doi.org/10.1109/access.2021.3057081).
- Oliveira, C. C.d., Löw, M.M. and Barros, T.H.B. (2024), “Knowledge organization possibilities for archives: comparative semantic analysis between cidoc-crm and ric-cm”, *KO Knowledge Organization*, Vol. 51 No. 5, pp. 362-370.
- Pires, C., Koch, I. and Nunes, S. (2023), “ArchOnto ontology representation of Portuguese archival description units (baptism records and passports)”, [Data set], doi: [10.25747/X78E-1A27](https://doi.org/10.25747/X78E-1A27).
- Python Software Foundation (2024a), “Datetime – basic date and time types”, available at: <https://docs.python.org/3/library/datetime.html> (accessed 20 December 2024).
- Python Software Foundation (2024b), “Re – regular expression operations”, available at: <https://docs.python.org/3/library/re.html> (accessed 20 December 2024).
- Richter, S. (2024), “Lxml – XML and HTML with Python”, available at: <https://lxml.de/> (accessed 20 December 2024).
- Riva, P., Zumer, M. and Aalberg, T. (2023), “LRMoo, navigating standards development processes in two communities”, *Committee on Standards Open Program International standards in the digital information landscape, World Library and Information Congress, 88th IFLA General Conference and Assembly*, Rotterdam, The Netherlands, available at: <https://repository.ifla.org/handle/123456789/2668>
- Rudwan, M.S.M. and Fonou-Dombeu, J.V. (2025), “A comparative analysis of fuzzy string matching algorithms for content-based ontology alignment”, in Arai, K. (Ed.), *Advances in Information and Communication*, Springer Nature Switzerland, Cham, pp. 141-156.
- Rupp, F., Schnabel, B. and Eckert, K. (2022), “Easy and complex: new perspectives for metadata modeling using RDF-star and named graphs”, in Villazón-Terrazas, B., Ortiz-Rodríguez, F., Tiwari, S., Sicilia, M.-A. and Martín-Moncunill, D. (Eds), *Knowledge Graphs and Semantic Web*, Springer International Publishing, pp. 246-262.

-
- Sikos, L. and Philp, D. (2020), "Provenance-aware knowledge representation: a survey of data models and contextualized knowledge graphs", *Data Science and Engineering*, Vol. 5 No. 3, pp. 293-316, doi: [10.1007/s41019-020-00118-0](https://doi.org/10.1007/s41019-020-00118-0).
- Tomasi, F. (2017), "Preserving cultural heritage objects: provenance formalization", *Bibliothecae.it*, Vol. 6 No. 2, pp. 17-40, available at: <https://bibliothecae.unibo.it/article/view/7531>
- Tomasi, F. (2023), "Archival finding aids in linked open data between description and interpretation", *JLIS.it*, Vol. 14 No. 3, pp. 134-146, doi: [10.36253/jlis.it-557](https://doi.org/10.36253/jlis.it-557).
- Varagnolo, D., Antas, G., Ramos, M., Amaral, S., Melo, D. and Rodrigues, I.P. (2022), "Evaluating and exploring text fields information extraction into CIDOC-CRM", *Proceedings of the 14th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2022) – KEOD*, SciTePress, pp. 177-184.
- Varagnolo, D., Melo, D., Rodrigues, I.P., Rodrigues, R. and Couto, P. (2023), "Archives metadata text information extraction into CIDOC-CRM", in *Knowledge Discovery, Knowledge Engineering and Knowledge Management: IC3K 2022*, Vol. 1842 *Communications in Computer and Information Science*, Springer, pp. 153-162.

(The Appendix follows overleaf)

Appendix

This appendix presents comparison tables of RDF code snippets that exemplify the scenarios presented in Chapter 3. The snippets encoded according to ArchOnto in Tables A1, A2, A3 are only partially aligned with Pires *et al.* (2023), as the model was updated. Consequently, modifications were made to the original RDF to reflect the current version of the model. The RDF snippets encoded in RiC-O in Tables A1, A2, A3 are excerpts from the dataset published in relation to the project (Giagnolini and Koch, 2024). As for Table A4, both snippets serve purely as illustrative examples and do not correspond to any entries in either dataset. Specifically, within the ArchOnto framework, data provenance coverage had not been previously considered until the discussions presented in this study. In the case of RiC-O representation, the depiction of data provenance was not deemed sufficiently expressive, leading to its omission from the final dataset.

Table A1. Comparison of ArchOnto and RiC-O representation of a birth event

| ArchOnto | RiC-O |
|--|--|
| <pre> <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:j.0="http://www.episa.inesctec.pt/archonto/#" xmlns:erlangen- crm="http://erlangen-crm.org/200717/" xmlns:j.1="http://www.episa.inesctec.pt/n-ary#" xmlns:j.2="http://www.episa.inesctec.pt/data_object#" xmlns:j.3="http://www.episa.inesctec.pt/ligacao#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:j.4="http://www.episa.inesctec.pt/isad-ontology#" xmlns:xsd="http://www.w3.org/2001/XMLSchema#" xmlns:owl="http://www.w3.org/2002/07/owl#" <owl:NamedIndividual rdf:about="http://erlangen-crm.org /200717/E31_Document761"> <erlangen-crm:P67_refers_to> <erlangen-crm:E67_Birth rdf:about="http://erlangen -crm.org/200717/E67_Birth3046"> <erlangen-crm:P98_was_born> <erlangen-crm:E21_Person rdf:about=" erlangen- crm.org/200717/E21_Person776 "> <erlangen-crm:P152_has_parent> <erlangen-crm:E21_Person rdf:about=" http://erlangen-crm.org /200717/E21_Person3049"/> </erlangen-crm:P152_has_parent> <erlangen-crm:P152_has_parent> <erlangen-crm:E21_Person rdf:about=" http://erlangen-crm.org /200717/E21_Person3050"/> </erlangen-crm:P152_has_parent> </erlangen-crm:E21_Person> </erlangen-crm:P98_was_born> <erlangen-crm:P97_by_father> <erlangen-crm:E21_Person rdf:about="http:// erlangen- crm.org/200717/E21_Person3049"> <erlangen-crm:P1_is_identified_by> <erlangen-crm:E41_Appellation rdf:about=" http://erlangen-crm.org/200717/ Manuel_Martins_Ramos"> <j.3:L2DO_hasValue> <j.2:DOE17_personName rdf:about="http ://erlangen-crm.org/200717/ Manuel_Martins_Ramos_"> <j.2:DOP5_name>Manuel Martins Ramos< /j.2:DOP5_name> </j.2:DOE17_personName> </j.3:L2DO_hasValue> </erlangen-crm:E41_Appellation> </erlangen-crm:P1_is_identified_by> </erlangen-crm:E21_Person> </erlangen-crm:P97_by_father> <erlangen-crm:P96_by_mother> <erlangen-crm:E21_Person rdf:about="http:// erlangen- crm.org/200717/E21_Person3050"> <erlangen-crm:P1_is_identified_by> <erlangen-crm:E41_Appellation rdf:about ="http://erlangen-crm.org/200717/ Ana_Joaquina_Martins"> <j.3:L2DO_hasValue> </pre> | <pre> <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:rico="https://www.w3.org/standards/RiC/ontology#" xmlns:xsd="http://www.w3.org/2001/XMLSchema"> <rico:RecordResource rdf:about="recordResource/PT-ADPRTPRQ-PPRT01- 001-0009-000096"> <rico:isAssociatedWithEvent> <rico:Event rdf:about="http://example.org/birth-of- Josefina-Martins-Ramos"> <rico:hasEventType rdf:resource="http://example.org/Birth"/> <rico:label>Birth of Josefina Martins Ramos</rico :label> <rico:occurredAtDate> <rico:Date rdf:about="http://example.org/date -1811-11-09"/> </rico:occurredAtDate> <rico:hasOrHadParticipant> <rico:person rdf:about="http://example.org/ person- Josefina-Martins-Ramos"> <rico:hasBirthDate> <rico:Date rdf:about="http://example.org/date- 1811-11-09"> <rico:expressedDate>1811-11-09</rico :expressedDate> <rico:normlizedDateValue>1811-11-09 T00:00:00</rico: normalizedDateValue> <rico:dateQualifier xsii:type="xsd: string">Certain </rico:dateQualifier> </rico:Date> </rico:hasBirthDate> <rico:isChildOf rdf:resources="http:// example.org/person-Manuel-Martins- Ramos"/> <rico:isChildOf rdf:resources="http:// example.org/person-Ana-JoaquinaMartins"/> </rico:person> </rico:hasOrHadParticipant> <rico:hasOrHadParticipant> <rico:person rdf:about="http://example.org/ person- Manuel-Martins-Ramos"> <rico:name>Manuel Martins Ramos</rico:name> </rico:person> </rico:hasOrHadParticipant> <rico:hasOrHadParticipant> <rico:person rdf:about="http://example.org/ person-Ana- Joaquina-Martins"> <rico:name>Ana Joaquina Martins</rico:name> </rico:person> </rico:hasOrHadParticipant> </rico:Event> </rico:isAssociatedWithEvent> <rico:isRelatedTo rdf:resource="http://example.org/ person-Josefina- Martins-Ramos"/> <rico:isRelatedTo rdf:resource="http://example.org/ person-Ana- Joaquina-Martins"/> </pre> |

(continued)

| | |
|--|--|
| <pre> <j.2:DOE17_personName rdf: about="http://erlangcrm.org/200717/ Ana_Joaquina_Martins"> <j.2:DOP5_name>Ana Joaquina Martins</j .2:DOP5_name> </j.2:DOE17_personName> </j.3:L2DO_hasValue> </erlangen-crm:E41_Appellation> </erlangen-crm:P1_is_identified_by> </erlangen-crm:E21_Person> </erlangen-crm:P96_by_mother> <erlangen-crm:P4_has_time-span> <erlangen-crm:E52_Time-Span rdf:about="http ://erlangen-crm.org/200717/E52_TimeSpan0002"> <erlangen-crm:P2_has_type> <j.4:ARE6_Date_Type rdf:about="http ://archonto/ARE6_Date_Type/ BirthDate/> </erlangen-crm:P2_has_type> <erlangen- crm:P1_is_identified_by> <erlangen-crm:E41_Appellation rdf: about="http://example.org /1811-11-09T00:00:00"> <j.3:L2DO_hasValue> <j.2:DOE10_instant> <j.2:DOP8_timestamp> 1818-11-09T00:00:00 </j.2: DOP8_timestamp> <erlangen-crm:P2_has_type > <j.0: ARE9_DateCertainty rdf:about="http ://www.episa. inesctec.pt/ archonto/# certain"/> </erlangen-crm: P2_has_type> </j.2:DOE10_instant> </j.3:L2DO_hasValue> </erlangen-crm:E41_Appellation> </erlangen-crm:P1_is_identified_by> </erlangen-crm:E52_Time-Span> </erlangen-crm:P4_has_time-span> </erlangen-crm:E67_Birth> </erlangen-crm:P67_refers_to> </owl:NamedIndividual> </rdf:RDF> </pre> | <pre> <rico:isRelatedTo rdf:resource="http://example.org/ person- Manuel-Martins-Ramos"/> </rico:RecordResource> </rdf:RDF> </pre> |
|--|--|

Source(s): Authors' own work

Table A2. Comparison of ArchOnto and RiC-O representation of the baptism activity

| ArchOnto | RiC-O |
|---|---|
| <pre> <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:j.0="http://www.episa.inesctec.pt/archonto/#" xmlns:erlangen- crm="http://erlangen-crm.org/200717/" xmlns:j.1="http://www.episa.inesctec.pt/n-ary#" xmlns:j.2="http://www.episa.inesctec.pt/data_object#" xmlns:j.3="http://www.episa.inesctec.pt/ligacao#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:j.4="http://www.episa.inesctec.pt/isad-ontology#" xmlns:xsd="http://www.w3.org/2001/XMLSchema#" xmlns:owl="http://www.w3.org/2002/07/owl#"> <owl:NamedIndividual rdf:about="http://erlangen-crm.org /200717/E31_Document761"> <erlangen-crm:P128_is_carried_by> <erlangen-crm:E22_Human-Made_Object rdf:about=" http://erlangen-crm.org/200717/ E22_humanmade_object761"> <erlangen-crm:P108_was_produced_by> <erlangen-crm:E12_Production rdf:about=" http://erlangen-crm.org/200717/ E12_Production761"> <erlangen-crm:P4_has_time-span> <erlangen-crm:E52_Time-Span rdf: about="http://erlangen-crm.org /200717/E52_Time-Span3057"> <erlangen-crm:P2_has_type> <j.4:ARE6_Date_Type rdf: about="http://archonto/ ARE6_Date_Type0001"/> <erlangen- crm:P2_has_type> <erlangen-crm: P1_is_identified_by> <erlangen- crm: E41_Appellation rdf: about="http://example. org/1811-11-16T00 :00:00"/> </erlangen-crm: P1_is_identified_by> </erlangen-crm:E52_Time-Span> <erlangen-crm:P4_has_time-span> </erlangen-crm:E12_Production> </erlangen-crm:P108_was_produced_by> </erlangen-crm:E22_Human-Made_Object> </erlangen-crm:P128_is_carried_by> <erlangen-crm:P70_documents> <erlangen-crm:E7_Activity rdf:about="http://erlangen- crm.org/200717/E7_Activity3043"> <erlangen-crm:P2_has_type rdf:resource="http://erlangen- crm.org/200717/Baptism"/> <erlangen-crm:P4_has_time-span> <erlangen-crm:E52_Time-Span rdf:about="http ://erlangen- crm.org/200717/E52_TimeSpan3057"> <erlangen-crm:P2_has_type> <j.4:ARE6_Date_Type rdf:about="http ://archonto/ARE6_Date_Type0002 "/> </erlangen-crm:P2_has_type> <erlangen-crm:P1_is_identified_by> <erlangen-crm:E41_Appellation rdf: about="rdf:about="http://example.org/1811-11-16T00 :00:00"/> </erlangen-crm:P1_is_identified_by> </erlangen-crm:E52_Time-Span> </erlangen-crm:P4_has_time-span> </erlangen-crm:E7_Activity> </erlangen-crm:P70_documents> </owl:NamedIndividual> <erlangen-crm:PC14_carried_out_by rdf:about="http://erlangen- crm.org/200717/PC14_carried_out_by3044"> <erlangen-crm:P01_has_domain> <erlangen-crm:E7_Activity rdf:about="http://erlangen- crm.org/200717/E7_Activity3043"/> </erlangen-crm:P01_has_domain> <erlangen-crm:P02_has_range> <erlangen-crm:E21_Person rdf:about="http://erlangen- crm.org/200717/E21_Person776"> </pre> | <pre> <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:rico="https://www.ica.org/standards/RiC/ontology#"> <rico:RecordResource rdf:about="recordResource/PT-ADPRTPRQ-PPRT01- 001-0004-000003"> <rico:documents> <rico:Activity rdf:about="http://example.org/baptof-Josefina- Martins-Ramos"> <rico:hasActivityType rdf:resource="http:// example.org/Baptism"/> <rico:label>Baptism of Josefina Martins Ramos</rico:label> <rico:occurredAtDate> <rico:Date rdf:about="http://example.org/date-1811- 11-16"> <rico:expressedDate> 11 de Novembro de 1811 </rico:expressedDate> <rico:normalizedDateValue>1811-11-16T00:00:00 </rico:normalizedDateValue> <rico:dateQualifier xsi:type="xsd:string">Certain </rico:dateQualifier> </rico:Date> </rico:occurredAtDate> <rico:thingsSourceOfRelation> <rico:PerformanceRelation rdf:about="http ://example.org/person-Josefina- Martins-Ramos-performed-as-baptized"> <rico:relationHasTarget> <rico:person rdf:resource="http:// example.org/person-JosefinaMartins- Ramos"> <rico:name>Josefina Martins Ramos </rico:name> </rico:person> </rico:relationHasTarget> <rico:relationHasContext> <rico:RoleType rdf:resource="http:// example.org/BaptizedPerson"/> </rico:relationHasContext> </rico:PerformanceRelation> </rico:thingsSourceOfRelation> <rico:thingsSourceOfRelation> <rico:PerformanceRelation rdf:about="http ://example.org/person-Josefina-Mariade-Jesus- performed-as-GodmotherOrGodfather"> <rico:relationHasTarget> <rico:person rdf:resource="http:// example.org/person-Josefina- Maria-de-Jesus"> <rico:name>Josefina Maria de Jesus </rico:name> </rico:person> </rico:relationHasTarget> <rico:relationHasContext> <rico:RoleType rdf:resource="http:// example.org/ GodmotherOrGodfather"/> </rico:relationHasContext> </rico:PerformanceRelation> </rico:thingsSourceOfRelation> </rico:Activity> </rico:documents> </rico:RecordResource> </rdf:RDF> </pre> |

(continued)

Table A2. Continued

| | |
|--|--|
| <pre><erlangen-crm:P1_is_identified_by> <erlangen-crm:E41_Appellation rdf:about="http:// erlangen- crm.org/200717/Josefina"> <j.3:L2DO_hasValue> <j.2:DOE17_person_name> <j.2:DOP5_name>Josefina Ramos</ j.2:DOP5_name> </j.2:DOE17_person_name> </j.3:L2DO_hasValue> </erlangen-crm:E41_Appellation> </erlangen-crm:P1_is_identified_by> </erlangen-crm:E21_Person> </erlangen-crm:P02_has_range> <erlangen-crm:P14.1_in_the_role_of> <j.0:P14.1_in_the_role_of rdf:resource="http:// erlangen- crm.org/200717/Baptized"/> </erlangen-crm:P14.1_in_the_role_of> </erlangen-crm:PC14_carried_out_by> <erlangen-crm:PC14_carried_out_by rdf:about="http:// erlangen- crm.org/200717/PC14_carried_out_by3056"> <erlangen-crm:P01_has_domain> <erlangen-crm:E7_Activity rdf:about="http:// erlangen- crm.org/200717/E7_Activity3043"/> </erlangen-crm:P01_has_domain> <erlangen-crm:P02_has_range> <erlangen-crm:E21_Person rdf:about="http:// erlangen- crm.org/200717/E21_Person3055"> <erlangen-crm:P1_is_identified_by> <erlangen-crm:E41_Appellation rdf:about=" http://erlangen-crm.org/200717/ Josefina_Maria_de_Jesus"> <j.3:L2DO_hasValue> <j.2:DOE17_person_name rdf:about=" http://dataobject/ DOE17_person_name/ JosefinaMariaDeJesus"> <j.2:DOP5_name>Josefina Maria de Jesus</j.2:DOP5_name> </j.2:DOE17_person_name> </j.3:L2DO_hasValue> </erlangen-crm:E41_Appellation> </erlangen-crm:P1_is_identified_by> </erlangen-crm:E21_Person> </erlangen-crm:P02_has_range> <erlangen-crm:P14.1_in_the_role_of> <j.0:P14.1_in_the_role_of rdf:resource="http:// erlangen- crm.org/200717/Godmother"/> </erlangen-crm:P14.1_in_the_role_of> </erlangen-crm:PC14_carried_out_by> </rdf:RDF></pre> | |
|--|--|

Source(s): Authors' own work

Table A3. Comparison of ArchOnto and RiC-O representation of an archival record

| ArchOnto | RiC-O |
|---|---|
| <pre> <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:j.0="http://www.episa.inesctec.pt/archonto/#" xmlns:erlangen- crm="http://erlangen-crm.org/200717/" xmlns:j.1="http://www.episa.inesctec.pt/n-ary#" xmlns:j.2="http://www.episa.inesctec.pt/data_object#" xmlns:j.3="http://www.episa.inesctec.pt/ligacao#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:j.4="http://www.episa.inesctec.pt/isad-ontology#" xmlns:xsd="http://www.w3.org/2001/XMLSchema#" xmlns:owl="http://www.w3.org/2002/07/owl#"> <erlangen-crm:E31_Document rdf:about="http://erlangen-crm .org/200717/E31_Document761"> <rdfs:label>Registo de batismo de Josefina</rdfs:label> <erlangen-crm:P1_is_identified_by> <erlangen-crm:E42_Identifier rdf:about="http:// erlangen- crm.org/200717/reference-code-000003"> <j.3:L2DO_hasValue> <j.2:DOP7_stringValue>000003</j.2:DOP7_stringValue > </j.3:L2DO_hasValue> <erlangen-crm:P2_has_type> <erlangen-crm:ARE5_Identifier_Type>ReferenceCode</ erlangen- crm:ARE5_Identifier_Type> </erlangen-crm:P2_has_type> </erlangen-crm:E42_Identifier> </erlangen-crm:P1_is_identified_by> <erlangen-crm:ARP12_has_level_of_description> <erlangen-crm:ARE1_Level_of_Description>File</erlangen -crm:ARE1_Level_of_Description> </erlangen-crm:ARP12_has_level_of_description> <erlangen-crm:P72_has_language> <erlangen-crm:E56_Language rdf:about="http://example. org/Portuguese"> <rdfs:label>Por (português)</rdfs:label> </erlangen-crm:E56_Language> </erlangen-crm:P72_has_language> <erlangen-crm:P94i_was_created_by> <erlangen-crm:E65_Creation rdf:about="http://erlangencrm.org/200717/E65_Creation781" > <erlangen-crm:P4_has_time-span> <erlangen-crm:E52_Time-Span rdf:about="http:// erlangen- crm.org/200717/E52_Time-Span3057"> <erlangen-crm:P1_is_identified_by> <erlangen-crm:E41_Appellation rdf:about="http ://example.org/1811-11-16T00:00:00"> </erlangen-crm:P1_is_identified_by> </erlangen-crm:E52_Time-Span> </erlangen-crm:P4_has_time-span> </erlangen-crm:E65_Creation> </erlangen-crm:P94i_was_created_by> <j.4:ISAD9_has_scope> Manuel Martins Ramos e Ana Joaquina Martins, Avós maternos: João Martins de Oliveira e Joana Francisca da Silva, Avós paternos: Manuel Martins Ramos e Custodia Maria da Costa, Padrinhos: Antonio Martins de Oliveira e Josefina Maria de Jesus, Data de nascimento: 9 de Novembro de 1811 </j.4:ISAD9_has_scope> <erlangen-crm:P128_carries> <erlangen-crm:E22_Human-Made_Object rdf:about="http:// erlangen-crm.org/200717/Human-MadeObject0001"> <erlangen-crm:P45_consists_of> <erlangen-crm:E57_Material>Paper</erlangen-crm: E57_Material> </erlangen-crm:P45_consists_of> </erlangen-crm:E22_Human-Made_Object> </erlangen-crm:P128_carries> <erlangen-crm:P43_has_dimension> <erlangen-crm:E54_Dimension> <erlangen-crm:P90_has_value>120x210</erlangen-crm: P90_has_value> <erlangen-crm:P91_has_unit>Millimeters</erlangen-crm :P91_has_unit> </erlangen-crm:E54_Dimension> </erlangen-crm:P43_has_dimension> <erlangen-crm:P70_documents> <erlangen-crm:E7_Activity rdf:about="http://erlangencrm.org/200717/E7_Activity3043"> <erlangen-crm:P2_has_type rdf:resource="http:// erlangen- </pre> | <pre> <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:rico="https://www.ica.org/standards/RiC/ontology#"> <rico:RecordResource rdf:about="http://example.org/RecordResource/PT- ADPRT-PRQ-PPRT01-001-0004-000003"> <rico:isOrWasIncludedIn rdf:resource="http://example. org/recordResource/PT-ADPRT-PRQ-PPRT01 -001-0004"/> <rdf:type rdf:resource="https://www.ica.org/standards/ RiC/ontology#Record"/> <rico:hasOrHadLanguage> <rico:Language rdf:about="http://example.org/ Portuguese"> <rdfs:label>Por (português)</rdfs:label> </rico:Language> </rico:hasOrHadLanguage> <rico:hasCreationDate> <rico:Date rdf:about="http://example.org/date -1811-11-16"/> </rico:hasCreationDate> <rico:hasOrHadIdentifier> <rico:Identifier rdf:about="http://example.org/ reference-code- 000003"> <rico:textualValue>000003</rico:textualValue> <rico:hasIdentifierType> <rico:IdentifierType rdf:about="http:// example.org/ReferenceCode"/> </rico:hasIdentifierType> </rico:Identifier> </rico:hasOrHadIdentifier> <rico:title xml:lang="pt">Registo de batismo de Josefina</rico:title> <rdfs:label xml:lang="pt">Registo de batismo de Josefina</rdfs:label> <rico:scopeAndContent rdf:parseType="Literal"> "Manuel Martins Ramos e Ana Joaquina Martins Avós maternos: João Martins de Oliveira e Joana Francisca da Silva Avós paternos: Manuel Martins Ramos e Custodia Maria da Costa Padrinhos: Antonio Martins de Oliveira e Josefina Maria de Jesus Data de nascimento: 9 de Novembro de 1811" </rico:scopeAndContent> <rico:hasInstantiation> <rico:Instantiation rdf:about="http://example.org/ instantiation/PT-ADPRT-PRQ-PPRT01 -001-0004-000003-11"> <rico:hasExtent> <rico:CarrierExtent rdf:about="http:// example.org/carrier-extent-120x210mm"> <rico:textualValue>120x210mm ; papel</ rico:textualValue> <rico:hasUnitOfMeasurement> <rico:UnitOfMeasurement rdf:resource ="http://example.org/ Millimeters"/> </rico:hasUnitOfMeasurement> </rico:CarrierExtent> </rico:hasExtent> <rico:hasCarrierType> <rico:CarrierType rdf:about="http://example .org/Papel"/> </rico:hasCarrierType> <rico:hasOrHadIdentifier> <rico:Identifier rdf:about="http://example. org/physical-location-E /20/6/3-9.4-fl.-3.0-ass.3"> <rico:textualValue>E/20/6/3 - 9.4 - fl. 3 v., ass.3</rico:textualValue> <rico:hasIdentifierType> <rico:IdentifierType rdf:about="http ://example.org/ PhysicalLocation"/> </rico:hasIdentifierType> </rico:Identifier> </rico:hasOrHadIdentifier> <rico:hasCreationDate> <rico:Date rdf:about="http://example.org/ date-1811-11- 16"/> </rico:hasCreationDate> <rico:title xml:lang="pt">Registo de batismo de Josefina </pre> |

(continued)

Table A3. Continued

| | |
|---|--|
| <pre> crm.org/200717/Baptism"/> </erlangen-crm:E7_Activity> </erlangen-crm:P70_documents> <erlangen-crm:P67_refers_to> <erlangen-crm:E67_Birth rdf:about="http://erlangen- -crm.org/200717/E67_Birth3046"> <erlangen-crm:P4_has_time-span> <erlangen-crm:E52_Time-Span> <erlangen-crm:P1_is_identified_by> <erlangen-crm:E41_Appellation rdf: about="http://example.org /1811-11-09T00:00:00"/> </erlangen-crm:P1_is_identified_by> <erlangen-crm:P2_has_type> <j.4:ARE6_Date_Type rdf:about="http ://example.org/BirthDate"/> </erlangen-crm:P2_has_type> </erlangen-crm:E52_Time-Span> </erlangen-crm:P4_has_time-span> </erlangen-crm:E67_Birth> </erlangen-crm:P67_refers_to> </erlangen-crm:E31_Document> </rdf:RDF> </pre> | <pre> </rico:title> </rico:Instantiation> </rico:hasInstantiation> <rico:documents> <rico:Activity rdf:about="http://example.org/baptof-Josefina- Martins-Ramos"> </rico:Activity> </rico:documents> <rico:isAssociatedWithEvent> <rico:Event rdf:about="http://example.org/birth-of -Josefina- Martins-Ramos"> </rico:Event> </rico:isAssociatedWithEvent> </rico:RecordResource> </rdf:RDF> </pre> |
|---|--|

Source(s): Authors' own work

Table A4. Comparison of ArchOnto and RiC-O representation of data provenance

| ArchOnto | RiC-O |
|---|---|
| <pre> <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:0="http://www.episa.inesctec.pt/archonto/#" xmlns:erlangen- crm="http://erlangen-crm.org/200717/" xmlns:j.1="http://www.episa.inesctec.pt/n-ary#" xmlns:j.2="http://www.episa.inesctec.pt/data_object#" xmlns:j.3="http://www.episa.inesctec.pt/ligacao#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:j.4="http://www.episa.inesctec.pt/isad-ontology#" xmlns:xsd="http://www.w3.org/2001/XMLSchema#"> <erlangen-crm:E31_Document rdf:about="http://example.org/ Document0001"> <erlangen-crm:P165_incorporates> <erlangen-crm:D1_Digital_Object rdf:about="http ://example.org/DigitalObject0001"/> </erlangen-crm:P165_incorporates> <erlangen-crm:P165_incorporates> <erlangen-crm:D1_Digital_Object rdf:about="http ://example.org/DigitalObject0002"/> </erlangen-crm:P165_incorporates> <erlangen-crm:E31_Document> <erlangen-crm:D1_Digital_Object rdf:about="http:// example.org/DigitalObject0001"> <erlangen-crm:P1_is_identified_by> <erlangen-crm:E41_Appellation rdf:about="http:// example.org/Appellation0006"> <j.3:L2DO_has_value> <j.2:DOE8_String rdf:about="http://example. org/String0001"> <j.2:DOP7_stringValue>PT-ADPRT-PRQ- PPRT01.xml</j.2:DOP7_stringValue> </j.2:DOE8_String> <j.3:L2DO_has_value> <erlangen-crm:E41_Appellation> </erlangen-crm:P1_is_identified_by> </erlangen-crm:D1_Digital_Object> <erlangen-crm:D1_Digital_Object rdf:about="http:// example.org/DigitalObject0002"> <erlangen-crm:P1_is_identified_by> <erlangen-crm:E41_Appellation rdf:about="http:// example.org/Appellation0007"> <j.3:L2DO_has_value> <j.2:DOE8_String rdf:about="http://example. org/String0002"> <j.2:DOP7_stringValue>PT-ADPRT-PRQ- PPRT01_edited.rdf</j.2: DOP7_stringValue> </j.2:DOE8_String> <j.3:L2DO_has_value> <erlangen-crm:E41_Appellation> </erlangen-crm:P1_is_identified_by> </erlangen-crm:D1_Digital_Object> <erlangen-crm:D7_Digital_Machine_Event rdf:about="http ://example.org/DigitalMachineEvent0001"> <erlangen-crm:P2_has_type> <j.4:ARE16_Event_Type>Data extraction and migration</j.4:ARE16_Event_Type> </erlangen-crm:P2_has_type> <erlangen-crm:P3_has_note>This activity has been carried out... </erlangen-crm:P3_has_note> <erlangen-crm:L23_used_software_or_firmware> <erlangen-crm:D14_Software rdf:about="http:// example.org/Software0001"/> </erlangen-crm:L23_used_software_or_firmware> <erlangen-crm:P4_has_time-span> <erlangen-crm:E52_Time-Span rdf:about="http:// example.org/Time-Span0001"/> </erlangen-crm:P4_has_time-span> <erlangen-crm:P4_has_time-span> <erlangen-crm:L10_had_input> <erlangen-crm:D1_Digital_Object rdf:about="http ://example.org/DigitalObject0001"/> </erlangen-crm:L10_had_input> <erlangen-crm:L11_had_output> <erlangen-crm:D1_Digital_Object rdf:about="http ://example.org/DigitalObject0002"/> </erlangen-crm:L11_had_output> </erlangen-crm:D7_Digital_Machine_Event> </pre> | <pre> <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns #" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:rico="https://www.ica.org/standards/RiC/ontology#"> <rico:Record rdf:about="http://example.org/Record0001"> <rico:hasOrHadInstantiation> <rico:Instantiation rdf:about="http://example.org/ Instantiation0002"> <rdfs:label>PT-ADPRT-PRQ-PPRT01.rdf</rdfs: label> </rico:Instantiation> </rico:hasOrHadInstantiation> <rico:hasDocumentaryFormType> <rico:DocumentaryFormType rdf:about="http:// example.org/FindingAid"/> </rico:hasDocumentaryFormType> <rico:describesOrDescribed> <rico:RecordSet rdf:about="http://example.org/ RecordSet0001"/> </rico:describesOrDescribed> </rico:Record> <rico:Record rdf:about="http://example.org/Record0002"> <rico:hasOrHadInstantiation> <rico:Instantiation rdf:about="http://example.org/ Instantiation0003"> <rdfs:label>PT-ADPRT-PRQ-PPRT01.xml</rdfs:label> </rico:Instantiation> </rico:hasOrHadInstantiation> <rico:hasDocumentaryFormType> <rico:DocumentaryFormType rdf:about="http:// example.org/FindingAid"/> </rico:hasDocumentaryFormType> </rico:Record> <rico:Activity rdf:about="http://example.org/Activity0002"> <rico:hasActivityType> <rico:ActivityType rdf:resource="http://example. org/DataExtractionConversion"/> </rico:hasActivityType> <rico:history rdf:datatype="http://www.w3.org/2001/ XMLSchema#string"> This activity has been carried out... </rico:history> <rico:hasOrHadParticipant> <rico:Mechanism rdf:about="http://example.org/ Mechanism0001"> <rico:name>RiC-O converter</rico:name> </rico:Mechanism> </rico:hasOrHadParticipant> <rico:hasOrHadParticipant> <rico:Person rdf:about="http://example.org/ Archivist0001"> <rico:name>Lucia Giagnolini</rico:name> <rico:isOrWasControllerOf> <rico:Mechanism rdf:resource="http:// example.org/Mechanism0001"/> </rico:isOrWasControllerOf> </rico:Person> </rico:hasOrHadParticipant> <rico:hasBeginningDate> <rico:Date rdf:about="http://example.org/Date0003"/> </rico:hasBeginningDate> <rico:hasEndDate> <rico:Date rdf:about="http://example.org/Date0004"/> </rico:hasEndDate> <rico:isOrganicOrFunctionalProvenanceOf> <rico:Record rdf:resource="http://example.org/ Record0001"/> </rico:isOrganicOrFunctionalProvenanceOf> </rico:Activity> <rico:DerivationRelation rdf:about="http://example.org/ DerivationRelation0001"> <rico:relationHasSource rdf:resource="http://example. org/Instantiation0003"/> <rico:relationHasTarget rdf:resource="http://example.org/ Instantiation0002"/> <rico:relationHasContext rdf:resource="http://example. </pre> |

(continued)

Table A4. Continued

| | |
|--|---|
| <pre> <erlangen-crm:D14_Software rdf:about="http://example.org/Software0001"> <erlangen-crm:P1_is_identified_by> <erlangen-crm:E41_Appellation rdf:about="http://example.org/Appellation0008"> <j.3:L2DO_has_value> <j.2:DOE8_String rdf:about="http://example.org/String0003"> <j.2:DOP7_stringValue>RIC-O converter</j.2:DOP7_stringValue> </j.2:DOE8_String> </j.3:L2DO_has_value> </erlangen-crm:E41_Appellation> </erlangen-crm:P1_is_identified_by> </erlangen-crm:D14_Software> <erlangen-crm:E52_Time-Span rdf:about="http://example.org/Time-Span0001"> <erlangen-crm:P1_is_identified_by> <erlangen-crm:E41_Appellation rdf:about="http://example.org/Appellation0001"> <j.3:L2DO_hasValue> <j.2:DOE11_Interval rdf:about="http://example.org/Interval0001"> <j.0:ARE9_DateCertainty rdf:about="http://www.episa.inesctec.pt/archont/#certain"/> <j.2:DOP6_start_date_value rdf:datatype="xsd:date">2024-07-01T00:00:00</j.2:DOP6_start_date_value> <j.2:DOP2_end_date_value rdf:datatype="xsd:date">2024-07-31T23:59:59</j.2:DOP2_end_date_value> </j.2:DOE11_Interval> </j.3:L2DO_hasValue> </erlangen-crm:E41_Appellation> </erlangen-crm:P1_is_identified_by> <erlangen-crm:P2_has_type> <j.4:ARE6_Date_Type rdf:about="http://example.org/DataMigrationDate"/> </erlangen-crm:P2_has_type> <erlangen-crm:E52_Time-Span> <erlangen-crm:PC14_carried_out_by rdf:about="http://example.org/CarriedOutBy0002"> <erlangen-crm:P14.1_in_the_role_of> <j.4:ARE8_Role_Type>Archivist</j.4:ARE8_Role_Type> </erlangen-crm:P14.1_in_the_role_of> <erlangen-crm:P02_has_range> <erlangen-crm:E21_Person rdf:about="http://example.org/Archivist0001"> <erlangen-crm:P1_is_identified_by> <erlangen-crm:E41_Appellation rdf:about="http://example.org/Appellation0005"> <j.3:L2DO_has_value> <j.2:DOE17_Person_Name rdf:about="http://example.org/PersonName0002"> <j.2:DOP5_name>Lucia Giagnolini</j.2:DOP5_name> </j.2:DOE17_Person_Name> </j.3:L2DO_has_value> </erlangen-crm:E41_Appellation> </erlangen-crm:P1_is_identified_by> </erlangen-crm:E21_Person> </erlangen-crm:P02_has_range> <erlangen-crm:P01_has_domain> <erlangen-crm:D7_Digital_Machine_Event rdf:about="http://example.org/DigitalMachineEvent0001"/> </erlangen-crm:P01_has_domain> </erlangen-crm:PC14_carried_out_by> </rdf.RDF> </pre> | <pre> org/Activity0002"/> </rico:DerivationRelation> </rdf.RDF> </pre> |
|--|---|

Source(s): Authors' own work

Corresponding authorLucia Giagnolini can be contacted at: lucia.giagnolini2@unibo.it

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgroupublishing.com/licensing/reprints.htmOr contact us for further details: permissions@emeraldinsight.com