ICA Records in Contexts-Ontology (RiC-O): a Semantic Framework for Describing Archival Resources*

Florence Clavaud[1] and Tobias Wildi[2]

[1] Archives nationales, Pierrefitte-sur-Seine, France
[2] University of Applied Sciences of the Grisons, Chur, Switzerland
florence.clavaud@culture.gouv.fr, tobias.wildi@fhgr.ch

Abstract. This article gives an overview of the new Records in Contexts Ontology (RiC-O), which is available at https://www.ica.org/standards/RiC/ontology. This ontology is part of the Records in Contexts (RiC) standard, which has been developed by the International Council on Archives to describe and contextualize archival resources in a comprehensible way that goes beyond the possibilities of the existing archival standards. The article explains the rationale for developing a new standard for archival description, and particularly the ontology. It provides a quick overview of the RiC Conceptual Model and then focuses on RiC-O, its design principles and content and giving references to more precise documentation that is publicly available online. Finally, it presents the roadmap and future perspectives of RiC.

Keywords: Archival metadata, Archival description, Records in Contexts, RiC, RiC-O, Linked Data, ontology, data modelling.

1 Introduction

This article gives an overview of the new Records in Contexts Ontology (RiC-O). This ontology is part of Records in Contexts (RiC) standard, that is being developed by EGAD,[1] a group of experts mandated by the International Council on Archives (ICA),[2]

[1] See: https://www.ica.org/en/about-egad (last accessed 2021/07/03). ICA has extended the initial mandate of EGAD to the end of 2021. This mandate will most probably be extended again.

[2] ICA is an international, neutral, non-governmental organization, which for more than sixty years "has united archival institutions and practitioners across the globe to advocate for good archival management and the physical protection of recorded heritage, to produce reputable standards and best practices, and to encourage dialogue heritage, to produce reputable standards and best practices, and to encourage dialogue, exchange, and transmission of this knowledge and expertise across national borders." EGAD is one of the expert groups established by the ICA Executive Board, on the recommendation of the Programme Commission. For more information, see https://www.ica.org/en/international-council-archives-0 and https://www.ica.org/en/our-professional-programme/expert-groups (last accessed 2021/07/03).

to describe and contextualize archival resources in a comprehensible way that goes beyond the possibilities of the previous archival standards.

The current version of RiC-O is version 0.2, that was published in February 2021.[3] It is an OWL 2 implementation of RiC-Conceptual Model (RiC-CM), the first part of RiC, whose full 0.2 version, including a long introduction and diagrams, was published in July 2021 [1] along with a call for comments.[4]

This article explains the rationale for developing a new standard for archival description, and particularly the ontology. It provides a quick overview of RiC-CM and then focuses on RiC-O, its design principles and content, giving references to more precise documentation that is publicly available online. Finally, it presents the roadmap and perspectives of the project.

2 Why Develop a New Standard for Archival Description?

Like the existing ICA standards for archival description,[5] RiC can be applied to any kind of archives, whatever their provenance, date, status, or nature is and whether they contain analogue or digital material. The perimeter of RiC is the same as it was for the previous standards: describing all kinds of archival material. However, RiC is very different from the existing ICA standards that it replaces.

The previous standards were neither pure conceptual models that would accurately define the domain described, nor technical, machine-readable implementations. They were something in between and aimed to specify description rules for exactly four categories of objects: archival resources, archival institutions, activities from which the archival resources result, and persons or groups related to archival resources. These four previous standards were also published independently between 1994 and 2008, which finally resulted in some overlapping and a few inconsistencies [2]. Besides, from 1998, the archival community (to be more precise, an international group led by the Society of American Archivists[6]) has developed and maintained, for authoring or, more often, exchanging archival metadata conforming to these description rules, two XML grammars, EAD [3] and EAC-CPF [4], that correspond to ISAD(G) and ISAAR(CPF) (and another DTD, EAG [5], that corresponds to ISDIAH and is far less used). So most of the huge quantity of archival metadata that is managed by archival institutions, at least for public historical archives, is available under the form of XML files that conform to two different XML grammars. These models together with the description rules mentioned above have shaped the landscape of archival metadata and enabled the

[3] See: https://www.ica.org/standards/RiC/ontology (last accessed 2021/09/01).

[4] See: https://www.ica.org/en/call-for-comments-on-ica-ric-cm-02 (last accessed 2021/09/01).

[5] ISAD(G), ISAAR(CPF), ISDF and ISDIAH; see: https://www.ica.org/en/public-resources/standards (last accessed 2021/07/03).

[6] The Technical Subcommittee on Archival Encoded Standards (TS-EAS), a subcommittee of the SAA Standards Committee. See: https://www2.archivists.org/groups/technical-subcommittee-on-encoded-archival-standards-ts-eas (last accessed 2021/09/01).

development of portals like Archives Portal Europe.[7] However, even if these description rules are considered essential by the archives, they are not based on a global, consistent and rigorous conceptual framework. The components of the standards as well as the corresponding elements and attributes in the three existing XML models do not have the appropriate granularity to cover the current needs of archivists or records managers. For instance, ISAD(G) description rules apply to sets or records and to unitary records the same way. Moreover, the relations that are listed in ISAAR(CPF), ISDF and ISDIAH are too few and too vague to express the complexity and nuances of the archival world.

RiC as a completely new developed standard takes into account the latest developments of archival theory and practice. It bundles the preexisting four different standards into only one and aims to provide everything needed for describing archives and to create more homogenous data.

The key statement of RiC can be summarized as follows: *archives are by essence aggregations of records; each record and aggregation has its history and is engaged in multiple layers of contexts, from its creation to the management by an archival institution*. Examples of contexts include: the provenance, both organic and functional; the agents that authored the records; the documentary context (the relations between the record resource described and other records within the same aggregation or outside of it); the things the records are about; the successive actions (enrichment and modification, transfer, destruction or removal, appraisal, arrangement, classification, description, digitization, migration…) that were applied through time to the record sets; or the events (wars, earthquakes, floods…) that affected them; etc.

No archival aggregation (record set) can be well curated (arranged, described), or used by end users without the knowledge of all these layers of contexts. However, ISAD(G) and EAD mainly focus on the close documentary context, using "levels of description" in ISAD(G) and the XML tree to represent it. This is a lack of precision of the existing standards. Another such example is the fact that ISAAR(CPF) only distinguishes three agent types and five generic relation types between them. The new RiC standard aims to fill such gaps. It defines for example Record Sets and Records, and an explicit 'includes or included' relation between them, as well as a sequence relation, an association relation, a genetic relation, etc. RiC also defines six agent entities, and a significant number of relations that may exist between them. Thus, RiC intends to generalize, extend and express the underlying perspective on archives and opens possibilities beyond the previous description rules and XML structures.

RiC also takes into account current ideas of data modelling. Archival metadata was for a long time encoded and stored in tables (Excel/CSV), relational databases or the markup language XML, where EAD is the most prominent representative. To describe records in their different contexts, a more advanced way of data modelling is needed. *The most appropriate method for representing such a social and documentary network and its changes through time and space is an oriented, labeled graph of linked entities or, in other terms, a semantic network*.

---

[7] About Archives Portal Europe, see: https://www.archivesportaleurope.net/about-us (last accessed 2021/09/01).

Current archival information systems consist most often of silos that were created through time and are now coexisting, from specific heterogeneous databases and EAD authoring and publishing tools to complete digital archiving information systems. Within these systems, the same contextual entity, e.g. a person, a place, an event, or even the core archival entities i.e. record resources, may be described many times in different silos, with no link between these descriptions. Some contextual entities are not described independently as autonomous objects, but are named repeatedly, using simple textual strings within finding aids, e.g. for some access or appraisal rules. This results in various problems in managing information, such as lack of efficiency, difficulty to know what data the archival institution holds exactly about the entity, thus what knowledge it has about it, quality management issues and others. When end users try to retrieve records, they run into problems like lack of accuracy and insufficient recall when using the search engines provided, interfaces and lists of results that are difficult to understand, links between items that are not visible nor processable when they exist, or even several interfaces provided for the same archival institution.

RiC provides a global reference framework for moving away from such systems to data-oriented architectures, or at least for breaking up these internal silos and making them permeable. Once the metadata are reorganized per entity according to RiC, each entity described can be assigned a unique identifier and linked to other entities by using one of RiC relations. The graph obtained can be published and connected to other graphs maintained by other cultural institutions or research teams, as the same person or place, for example, can be of interest in another context. A record resource held in one institution can be directly linked to a resource in another archive. RiC is intended to be the foundation for creating interoperable archival metadata, that have the potential to be compliant with the FAIR principles.[8]

Interoperability is not only important between the repositories within an institution, but also across institutions and projects. Archival institutions or portals and their audiences will undoubtedly gain from setting up the conditions for interoperability at the level of metadata, and therefore from using RDF for publishing high quality metadata, conforming to a unique, homogeneous data structure. There are many examples of collections dispersed between several institutions or several countries; at the scale of archival series of files, the administrative and business processes also lead to a distribution of documents between the different actors, and whether it is a question of better evaluating these archives or identifying the most complementary sources, linked metadata may be of major interest. Not to mention that describing the huge quantity of archival contextual entities like agents, places, and functions, cannot be done without reusing the datasets prepared by other professional communities, that could in turn find new information in the knowledge graphs built by archivists. Finally, projects involving both researchers in humanities (or communities of users like associations or genealogists) and archivists or record managers are becoming more numerous, and being able

[8] The FAIR Guiding Principles for scientific data management and stewardship (2016) are recommendations to improve "the Findability, the Accessibility, the Interoperability and the Reuse of digital assets", around which a community and various initiatives have developed. See the website: https://www.go-fair.org/ (last accessed 2021/09/01).

to provide these projects with accurately modeled metadata becomes important. RiC will play a major role in building such interoperable sets.

From the perspective of archive users of any kind (both professionals and the general public, including researchers and citizens), organizing descriptive metadata as a graph of interconnected entities is a great help, since every component of a graph can be queried and displayed to fulfill the needs to *find*, *identify*, *select*, *obtain*, and *explore* the records described. These five verbs are borrowed from the Library Reference Model (LRM) [6], that was developed by and for the libraries community, following the very same idea to model the world of edited items as a graph of entities. The verb *understand* should in our opinion be added to the user tasks for archival metadata, as a user needs to understand the layers of contexts for selecting the records that are relevant for his/her research.[9]

With this target in mind, EGAD decided to build both an abstract conceptual model for archival description, and an OWL implementation of it, so that high quality RDF graphs can be created by the archival community, published and linked to other graphs.

3 History of the Project

EGAD had its first meeting in November 2013. The group released a first draft version of RiC-CM (RiC-CM 0.1) in August 2016 and gave a first presentation of this version at the international conference on archives held in Seoul (South Korea) in September 2016.[10] A call for comments was issued at the same time; it was closed in January 2017 and resulted in a significant quantity of comments from more than sixty professional organizations and persons, which were of great help for preparing RiC-CM 0.2. Meanwhile, early versions of RiC-O were developed, which also proved to be useful for designing the conceptual model; they were thoroughly tested by EGAD[11] and also shared with early reviewers that had applied as volunteers.[12] The first public release of RiC-O occurred in December 2019, at the same time as the release of RiC-CM 0.2

---

[9] Let us also, of course, add the *reuse* verb, as data, and particularly RDF metadata sets, are designed to be reused outside of their original environment, particularly by researchers, and by machines.

[10] See the slides and audio recording of the presentation: https://www.ica.org/en/records-in-contexts-ric-a-standard-for-archival-description-presentation-congress-2016 (last accessed 2021/07/03).

[11] In particular, in February 2018, three French institutions (the Archives nationales of France, the French national Library and the French ministry of Culture) published a proof of concept, the PIAAF prototype (https://piaaf.demo.logilab.fr), based on a version of RiC-O dated 2017.

[12] The comments received in 2019 on early versions of RiC-O sent to the early reviewers consisted in seven PDF files. The authors were at this date working in archival institutions or service and software providers, that were located in Belgium, Bulgaria, Canada, France, Portugal, and the USA.

Preview.[13] Since then, the current version of RiC-O has therefore been accessible both to machines and humans through its IRI, https://www.ica.org/standards/RiC/ontology.

Both RiC-CM 0.2 Preview and RiC-O 0.1 had to be carefully checked and slightly modified, and RiC-CM 0.2 Preview lacked an introduction. As a result of this work, RiC-O 0.2 was published in February 2021; it is compliant with RiC-CM 0.2 full draft, which was published in July 2021 [1], along with a call for comments. Both are significantly different from the previous public releases, and supersede them. The list of changes made to RiC-O from December 2019 is included in the internal documentation of the OWL file.

From March 2020, the versioned sources of RiC-O have been available through the GitHub public repository of the project,[14] which makes it possible for any person to easily contact the RiC-O development team,[15] to create issues, to fork the repository and more generally speaking to follow up the development process. RiC-O also has an information website,[16] that includes pages on the RiC-O roadmap, on events and on projects using RiC-O. Let us also mention that RiC-O is also available in LOV.[17]

It is important to note that RiC-O 0.2, and RiC-CM 0.2, are still draft versions. EGAD plans to release the first stable version of RiC, having the status of an ICA recommendation, by the beginning of 2022. The comments that will be submitted to EGAD from now to the fall of 2021 will be incorporated into the standard and are therefore of high importance to EGAD.

4 Developing a Domain Standard

When the EGAD took up its work, no generic model or ontology for describing the domain of archival resources existed.[18] Various standards and models were analyzed and reviewed:

---

[13] Records in Contexts-Conceptual model (RiC-CM) 0.2 preview (December 2019), https://www.ica.org/sites/default/files/ric-cm-0.2_preview.pdf, last accessed 2021/09/01.

[14] See https://github.com/ICA-EGAD/RiC-O. The repository also already includes several full examples as well as diagrams. The current official version of RiC-O is to be found in the main branch, while the development of the next version is always being made using the 'next-version' branch.

[15] The most active members of RiC-O development team are the two authors of this article. Daniel Pitti, chair of EGAD, is also involved in this task.

[16] https://ica-egad.github.io/RiC-O/.

[17] https://lov.linkeddata.es/dataset/lov/vocabs/rico (last accessed 2021/07/03).

[18] Some national conceptual models already existed, like the Spanish NEDA conceptual model [13]. Some ontologies existed, developed by smaller communities for specific projects. Among them, let us quote the British, Jisc-funded, LOCAH project that lasted from 2010 to 2013 (http://locah.archiveshub.ac.uk/) and that resulted in datasets and a RDFS vocabulary (see http://data.archiveshub.ac.uk/); and Matterhorn RDF, which was a project to convert METS-based information packages to RDF (https://matterhorn.tools/).

- Among very generic ontologies or broader vocabularies: as an example, the widely used schema.org RDFS vocabulary is not precise and rich enough for enabling a complete archival description.[19]
- Among models for cultural heritage: FRBRoo [7] ontology or the IFLA LRM model [6] and RDA registry [8] do not fit for describing the vast majority of records, as records most often are not edited creative works; thus, the core LRM Work, Manifestation, Expression and Item entities are not relevant. Also, records are not the primary focus of CIDOC CRM [9], which, for example, lacks methods for handling aggregations of records and the relations between them through time, and lacks a lot of datatype properties for describing archival resources.[20] There are important differences between describing general cultural heritage objects and archival records, which is one reason why CIDOC CRM is not suitable for modelling archival content.
- Both the PREMIS [10] as well as the EBUCore ontology [11] focus on technical and administrative metadata, in particular to ensure the long-term preservation of digital records; though preservation and descriptive metadata overlap, standards for technical metadata cannot just be reused for the description of entities.
- The PROV Ontology [12], which is a domain-agnostic ontology to model the provenance of data, is a good starting point, but too generic to be able to represent the specific requirements of archives for the modeling of traceability and provenance.

An intensive analysis led EGAD to the decision to develop its own conceptual model with a domain ontology for archives. The members of the EGAD had not only to work on the necessary definitions, but also to include the various feedbacks from their peers and institutions in the archival world. Once RiC is released in its final form, the long term maintenance will be assured by the ICA on behalf of, and with, the archival community. New entities, attributes and relations will have to be added, old ones to be clarified or changed.

5 Quick Overview of RiC-CM 0.2

RiC-CM consists of 22 entities, among which four are the core ones. The core entities are interconnected with each other: Record Resource and the closely related Instantiation entity, as well as the Agent and Activity entities. These four entities are considered essential in describing records and the contexts in which these records emerged and were used over time. Broadly speaking, the four core entities and their relations model

[19] Schema.org is primarily being developed for including "structured data on the internet", that describes various kinds of resources. It includes an ArchiveComponent class (https://schema.org/ArchiveComponent), subclass of CreativeWork (https://schema.org/CreativeWork), whose properties, particularly object properties, are far from enabling an accurate description of records.

[20] For example, CRM only provides 'P106_is_composed_of' object property for connecting two instances of any E90_Symbolic_Object, while RiC provides 'includes or included' for connecting a Record Set to a Record Set or a Record that it includes or included during some period.

agents who act in time and space, fulfill administrative or other tasks and thus create records.

The core entities are not new for archivists and align closely with the existing ICA standards: ISAD(G) -- Record Resource, ISAAR(CPF) -- Agent, and ISDF -- Activity.[21] Among the core entities, only Instantiation is new. This entity helps to model different materializations of records and corresponds to the PREMIS Representation.

RiC-CM would not be a complete conceptual framework without a set of entities to model archival context further. Thus, apart from the Agent entities, RiC-CM also defines Event and Activity; Rule and Mandate; Place; Date.

RiC-CM defines 41 attributes for these entities, some of them being shared by all, some being specific to a group of entities, and some to be used only with one entity (such as 'Scope and content', an attribute of only Record Resource). It provides a rich set of 78 relations (and the inverse ones when they are not symmetric) for linking the entities to each other; for example, 'has Instantiation' which connects a Record Resource and one of its Instantiations, or 'documents' which can connect a Record Resource or Instantiation and an Activity from which it results.

Let us finally mention that RiC-CM also provides sub-structures for some of the entities. For example, Record, Record Set and Record Part are sub-entities of Record Resource; Activity is a sub-entity of Event; Agent has in total six sub-entities (Person, Group that includes Family and Corporate Body, Position and Mechanism). Among Agent sub-entities, Position and Mechanism are rather specific to the archival domain compared with the models of other cultural heritage communities (libraries and museums).

This feature provides great flexibility for implementation, that which archivists and records managers have always absolutely needed when describing archives, because they will do it more or less precisely according to the nature of the record sets or items to be described, the knowledge that they have been able to accumulate on these archives, the means and the time available. So, when you don't know whether you are dealing with a Record or a Record Part or a Record Set (as can be the case with only reading many old inventories), you can use the Record Resource super-entity; conversely, the Record entity has its own attributes and is the domain of some specific relations, but it also inherits attributes from Record Resource and can be used with relations whose domain is Record Resource.

This feature also facilitates linking: high-level, fairly generic entities and relations, such as Group, Event or 'is or was part of', are good candidates for mappings with entities or relations from other models like PREMIS, PROV-O, IFLA LRM or CIDOC CRM, which will be done in future versions of RiC.

Finally, this poly-hierarchical organization facilitates extensions of the model to meet specific needs. Some attributes, moreover, are defined as extensible. If desired, one can define a relation which does not exist in RiC, but which would be needed locally, as a sub-relation of a RiC relation; it will then inherit its characteristics, domain, scope, and can always be replaced by the super-relation defined in RiC.

---

[21] See RiC-CM 0.2 [1], Introduction, section 1.6 (Relationships between RiC-CM and other models and standards for describing records), p. 7-13.

Finally, of course, this feature makes it easier to transpose RiC-CM into an OWL ontology.

The diagram below shows the most important entities of RiC and their relations; it can also be downloaded from the RiC-O Git repository.[22]

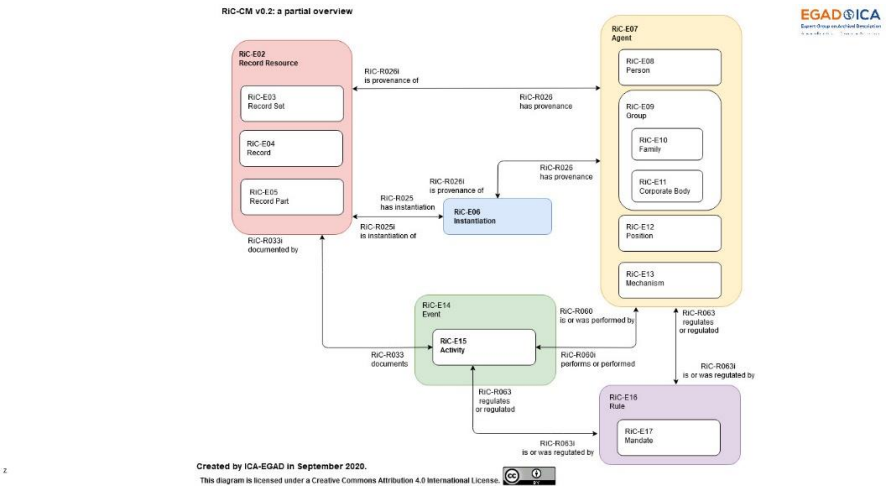

Fig. 1. Overview of RiC-CM main entities

6 RiC-O Design Principles and Main Characteristics

As explained in more details in the introduction of RiC-O,[23] the members of EGAD who developed RiC-O followed some design principles, among which we will present the most important ones.

6.1 RiC-O Has to Be Fully Compliant with RiC-CM, of which it Is the Official OWL 2 Representation

This means that RiC-O is a domain ontology, having the same scope as RiC-CM. It is developed in close interaction with the archival community. If needed, it will be updated and maintained by this very same community.

This also means that every component of RiC-CM must have a corresponding component in RiC-O. The RiC-O component has the same textual definition (rdfs:comment,

[22] https://raw.githubusercontent.com/ICA-EGAD/RiC-O/master/diagrams/diagrams_v0-2/RiC-CM-overview/diagram_RiC-CM-overview-RiC-v0-2.jpg.

[23] RiC-O design principles and a quick overview, including what distinguishes the ontology from the model, are presented in the OWL source file, and can be read as the introduction to RiC-O HTML view (see https://www.ica.org/standards/RiC/RiC-O_v0-2.html#design-principles, and https://www.ica.org/standards/RiC/RiC-O_v0-2.html#understanding-RiCO).

skos:scopeNote) as the RiC-CM component when there is a simple equivalence. The hierarchies of components must also be the same when applicable.

Therefore, every entity defined in RiC-CM corresponds to a class in RiC-O. Any RiC-CM attribute that has text as schema value, corresponds to a datatype property in RiC-O (for example, the 'scopeAndContent' and 'authenticityNote' of a Record). Any relation in RiC-CM corresponds to an object property in RiC-O. These components have the same definition, the same hierarchical position, and the same domain and range when applicable, in both parts of the standard.

RiC-CM attributes that have controlled value correspond to classes in RiC-O, which results in a hierarchy of classes as concerns the *type attributes (for example, the Documentary Form Type of a Record, or the Carrier Type of an Instantiation). Moreover, classes or properties have been added in RiC-O, that have no equivalence in RiC-CM. Examples are the Physical Location and the Coordinates classes, that constitute, with the Place class, and their properties, a way to accurately describe a geo-historical entity.[24] Another example is the Extent class that was added recently.

6.2 RiC-O Has to Be Immediately Usable

Usability is considered a key feature, without which RiC will not be tested and adopted widely in the next years. In particular, it is very important that existing archival metadata, that are created or generated in current archival information systems, can be converted to RDF conforming to RiC-O and pushed to the web of data (where very few archival metadata are available for now). This process must be possible without losing any data, structural or partially implicit information. What is at stake here is that metadata conforming to the previous existing ICA standards can be processed successfully.

In order to guarantee that RiC-O conforms to this principle, during the ongoing development process, a lot of successful testing has been made, using XML/EAD finding aids and XML/EAC-CPF authority records, that have been converted to RDF datasets, either by hand or using scripts. While some existing metadata sets may have a very fine level of granularity and accuracy, and might store controlled vocabularies, or describing curation events separately, a lot of metadata don't have the very precise structure that RiC-CM recommends. Even then, a conversion process is possible.

The usability of a data model also depends on its documentation. The current official release of RiC-O is fully documented in English, and a human-readable view in HTML has been published online. In the near future, this documentation will be translated into French and Spanish, the two other official languages of the ICA. RiC-O is also already

[24] In RiC-O, a place is considered both a geographical and historical entity. As a historical entity, among other features, it has a history, and may be preceded or succeeded by other Places. A Place also may have zero to many Physical Location through time (for instance, its boundaries, if it is an administrative area or a country, may change). Each Physical Location may be connected to zero to many Coordinates. This model is quite close to the Linked Places Format (https://github.com/LinkedPasts/linked-places).

accompanied with examples (RDF datasets) and diagrams,[25] which will be enriched through time. Some tutorials should also be written, and EGAD will organize practical workshops.

6.3 RiC-O Has to Provide a Flexible Framework

This principle is related to the usability principle quoted above. Archival description is flexible by essence. It is quite common that within the same archive the level of granularity of information varies from one finding aid to another (or from one authority record to another), or even within the same finding aid. Some series or agents are described summarily because little is known about them and there is little time for extensive research, while other series, even records, or agents are described in detail; some relations (e.g. the link to the provenance) may be described without any detail while others may be thoroughly documented, as ISAAR(CPF) and EAC-CPF allow it.

Being generally flexible, for an OWL ontology, depends first on the polyhierarchical systems of classes and properties it provides. A superproperty or superclass is normally very generic, its underlying subproperties and subclasses then become more concrete for handling detailed information. RiC-O is based on this principle. It lets the archivist work with a certain subset of RiC-O components, where, for example, high level classes or properties, appropriate for handling vague information on agents, can coexist with other more precise ones, necessary in a specific project to describe very accurately a collection of photographs or of mediaeval charters for example.

Besides, and this is related to the two 'immediate usability' and 'flexibility' design principles, RiC-O sometimes provides more than one way for representing information. RiC-CM, for example, defines a Date entity. RiC-O provides on the one hand a Date class with several subclasses, on the other hand a 'date' datatype property, that lets users encode dates in a simpler way, just as they were used to doing it in EAD or EAC-CPF files. Users then can choose from one of the methods, depending on which they want to describe a date as a full entity possibly with additional information, or just as one simple attribute among others.

RiC-O also deals with many relations between entities in a similar way: they are both represented as a binary object property and as a class, which is a subclass of the Relation class. Using a class for a relation allows adding additional information to a relation, like a date or a description. In the past, EAC-CPF has already modelled relations in a similar way. The direct object property is then defined as a shortcut of this more complex path, using OWL 2 property chain axiom.[26]

---

[25] Complete examples are available at https://github.com/ICA-EGAD/RiC-O/tree/master/examples/examples_v0-2, and diagrams at https://github.com/ICA-EGAD/RiC-O/tree/master/diagrams/diagrams_v0-2.

[26] See https://www.w3.org/TR/owl2-new-features/#F8:_Property_Chain_Inclusion (last accessed 2021/09/01). An example of such a relation is available through a diagram in RiC-O repository on GitHub, see: https://github.com/ICA-EGAD/RiC-O/blob/master/diagrams/diagrams_v0-2/NationalArchivesOfFrance-examples/a-complex-relation-and-its-shortcut.jpg.

6.4 RiC-O Must Open New Potential for Archival Description and Discovery

This principle takes up the interests and requirements of the end users. Linked Data tools and interfaces should enable users to go through RDF/RiC-O graphs, to query them using SPARQL in an efficient way and to consult archival metadata and their contexts in new ways. With the current archival information systems, a lot of such queries are not possible. As an example, users should be able to ask « What are (according to your dataset) the corporate bodies that succeeded to this given entity from its end of existence, by 1840, to nowadays (as concerns this given activity)? » or « Tell me what instantiations of this photograph exist? », or « What are the existing copies of this original charter? », and get a result list of the entities. In other words, institutions or projects that make the effort to implement RiC-O must get insights into the content and context of their archives that weren't possible with the existing ICA standards. It should be even more interesting if you can infer new assertions from the RDF datasets you built, and of course link your datasets to resources outside of your institution. This important feature is currently investigated in several projects.[27]

6.5 RiC-O Should Be Extensible

Institutions are free to extend the ontology by adding new subclasses or subproperties if needed. RiC-O has also the potential to be usable in other contexts than purely archival ones. This implies that hierarchies of classes and properties are defined and that mappings are developed with other ontologies as mentioned above. It also implies that RiC-O provides "hooks" enabling connections with, for example, existing SKOS vocabularies – which has been made possible by defining a hierarchy of Type classes, whose instances can easily be declared as also being SKOS concepts.[28]

7 Conclusion: Roadmap and Perspectives

The principles and choices presented above result in an OWL domain model that relays on a strong conceptual basis (RiC-CM). RiC-O as an implementation is richer and more accurate than its conceptual model, and at the same time grants all flexibility. RiC-O 0.2 (and the future 1.0 version) provides in many cases more than one method for representing information; from this point of view, it may be considered a transitional

---

[27] See for example a presentation by Florence Clavaud, "Implementing ICA Records in Contexts-Ontology at the National Archives of France: first assessment and prospects", for the Study Day on The Semantic Web and Cultural Heritage: From Data Convergence to Knowledge Crossing (Lille, France, February 3, 2021) - slides in English and audio recording in French: https://geriico.univ-lille.fr/detail-event/le-web-semantique-et-le-patrimoine-culturel-de-la-convergence-des-donnees-au-croisement-des-connai/ (last accessed 2021/07/03).

[28] As done by the Archives nationales of France, which have started to publish their vocabularies and authority records, using RiC-O and SKOS, in a GitHub public repository; see: https://github.com/ArchivesNationalesFR/Referentiels.

ontology, in which some components may be deprecated later on. In most cases, RiC-O will only partially be implemented and institutions will choose the elements that are useful for their local project, and simultaneously, might have to extend it.

As a conclusion, we would like to insist on the fact that RiC-CM and RiC-O are still at an early stage and that this is only the beginning of the adventure. After a certain time, as with any domain model and ontology under construction, various institutions will have made their first experiences with adopting RiC-O -- and adapting it to their needs. The new ontology provides in many cases more than one option to model a characteristic or a relation; RiC-O users will have to decide in which way they will take to model their use case. The intent of EGAD is to collect practical examples of how RiC-O is put in use and to identify some "best practices" that stand out and prevail.

Besides, the roadmap of EGAD foresees alignments with other ontologies, particularly the ones we quoted in section 4. In our opinion good mappings can only be achieved by working with the communities that authored these other models.

The further work might also include the development of SHACL Shapes[29] for the RiC-O classes and properties, in order to provide archival institutions and projects with methods for validating RiC-O knowledge graphs, at least cardinality, domain and range checks.

RiC is already being implemented in various projects,[30] among which many aim to publish RDF/RiC-O graphs, or already do so. The number of projects will hopefully rapidly grow. Thus, various kinds of tools will be needed and developed -- and hopefully published under licenses that makes reuse possible for other institutions.[31] By proposing a multidimensional, novel, consistent, inclusive insight into the content of archives, new perspectives open up and new territories will be explored.

In order to facilitate the construction of such linked archival data beyond institutional and national boundaries, another task for archivists and the ICA will be the development of multilingual and international authority data and vocabularies, based on SKOS and RiC-O, that could be used for specifying the categories of, for instance, documents or groups of documents, functions and occupations, curation events, access rules – in other words, what RiC calls archival contextual entities.

References

1. International Council on Archives (ICA): Records in Contexts-Conceptual model (RiC-CM) 0.2 (July 2021), https://www.ica.org/sites/default/files/ric-cm-02_july2021_0.pdf, last accessed 2021/09/01.
2. International Council on Archives (ICA) Committee on Best Practices and Standards: Progress report for revising and harmonising ICA descriptive standards (2012),

[29] These would use Shapes Constraint Language, see https://www.w3.org/TR/shacl/ (last accessed 2021/09/01).

[30] See https://ica-egad.github.io/RiC-O/projects-and-tools.html.

[31] Like RiC-O Converter (https://github.com/ArchivesNationalesFR/rico-converter), an open source tool that is presented in another article authored by Thomas Francart, Florence Clavaud, and Pauline Charbonnier.

https://www.ica.org/sites/default/files/Report_for_revising_harmonising_description_standards.pdf, last accessed 2021/09/01.
3. Encoded Archival Description (EAD): https://www.loc.gov/ead/, last accessed 2021/07/03.
4. Encoded Archival Context-Corporate Bodies, Persons, and Families (EAC-CPF) XML schema: https://eac.staatsbibliothek-berlin.de/, last accessed 2021/07/03.
5. Encoded Archival Guide (EAG): http://www.archivesportaleurope.net/eag, last accessed 2021/09/01.
6. International Federation of Libraries Associations and Institutions (IFLA): Library Reference Model (LRM) (2017), https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf, last accessed 2021/07/03.
7. International Working Group on FRBR and CIDOC CRM Harmonisation: Functional Requirements for Bibliographic Records-object oriented (FRBRoo) model version 2.4 (November 2015), http://iflastandards.info/ns/fr/frbr/frbroo/, last accessed 2021/07/03.
8. Resource Description and Access (RDA) Steering Committee: RDA Registry, https://www.rdaregistry.info/, last accessed 2021/07/03.
9. International Council of Museums (ICOM) International Committee for Documentation (CIDOC): CIDOC Conceptual Reference Model (CIDOC CRM) version 7.1.1 (April 2021), http://www.cidoc-crm.org/sites/default/files/cidoc_crm_v.7.1.1_0.pdf, last accessed 2021/07/03.
10. PREMIS (PREservation Metadata Implementation Strategies) Editorial Committee: PREMIS 3 ontology for PREMIS Data Dictionary version 3, http://www.loc.gov/premis/rdf/v3/, last accessed 2021/07/03.
11. European Broadcasting Union (EBU): EBUCore - The Dublin Core for media, https://www.ebu.ch/metadata/ontologies/ebucore/, last accessed 2021/07/03.
12. The World Wide Web Consortium (W3C): The PROV ontology (PROV-O) (2013), http://www.w3.org/TR/prov-o/, last accessed 2021/07/03.
13. Ministerio de Educación, Cultura y Deporte, Comisión de Normas Españolas de Descripción Archivística (CNEDA): Modelo conceptual de descripción archivística (2017), https://www.culturaydeporte.gob.es/cultura/areas/archivos/mc/cneda/documentacion/normas/neda-mc.html, last accessed 2021/09/01.