ICON PROJECT: CONTENT INTEGRATION IN PORTUGUESE NATIONAL ARCHIVES USING CIDOC-CRM

Maria José de Almeida[1], Lucília Runa[1]

[1] DGLAB, Alameda da Universidade, Edifício da Torre do Tombo, Lisboa, Portugal

m-jose.almeida@dglab.gov.pt

Abstract. DGLAB is responsible for the management of several information systems that support its mission to safeguard, enhance and promote governmental and public records, as well as other historical documents in its custody. One of the most significant information system (http://digitarq.arquivos.pt) holds the archival description of information objects dating from the 9th century to present days, of several natures and media. The data infrastructure is supported by a relational database, even though the archival descriptions obey the hierarchic rules defined by the standards of the International Council on Archives (ICA). Both the technology and the data model appear to be unable to respond to the current challenges of our information management needs and those of our audience.

The ICON Project aims to renew the existing data infrastructure in order to improve efficiency to internal and external users. One of its key features is content integration, as we intend to create a more flexible data model that can both interoperate with other information systems and accommodate information regarding cultural resources other than archival documents. As the ICA is reviewing its standards towards a conceptual model, but as it has not yet released a final or stable version, the choice of CIDOC-CRM as our root ontology seems the most appropriate for our project.

Focusing on the archival resources of the DGLAB, this paper presents the ICON Project current state of development, discussing our doubts and options on applying the CIDOC-CRM to archival resources in integration with other cultural objects information.

Keywords: Content Integration, Archival Resources, CIDOC-CRM, Portugal

1 Introduction

The *Direção Geral do Livro, dos Arquivos e da Biblioteca* – DGLAB (Directorate-General for Books, Archives and Libraries) is a central service under direct administration of the Portuguese State. Its mission is to coordinate the national archival system, regardless of each archive's form and medium, and its attributions are to structure, promote and monitor the intervention of the State in the definition of archival policy, to administer the appropriate measures for policy implementation, as well as the regime of protection of cultural heritage, promoting the safeguarding, enhancement, dissemination, access and enjoyment of archival heritage, its use as a resource for administrative activity and the foundation of collective and individual memory, guaranteeing the rights of the State and citizens ((Decreto-Lei 102/2012 2012). It is structured in nuclear organic units, among which the *Direção de Serviços de Inovação e Administração Eletrónica* – DSIAE (Innovation and Electronic Administration Services Directorate), which is responsible for designing and developing cross-sectional projects in functional areas of archiving, application of new technologies, and administrative modernisation; and of managing and bettering the national network of archives, including the development of information and communication structures aimed at maintaining and expanding the services (Portaria n.º 192/2012 2012). This is the context of ICON project (*Integração de CONteúdos*), started by DGLAB / DSIAE in 2017. ICON involves the 19 national and regional archives that currently constitute the DGLAB network. Its goal is to renew the architecture of its current information systems, which requires the definition of a new basic conceptual model. The existing systems are:

a) *DigitArq-CRAV:* supports records' archival description, document conservation and preservation actions, reading room orders; digital reproductions requests; certificates requests, etc.;
b) *Portal Português de Arquivos – PPA* (Portuguese Archives Portal): makes available, through a single centralised point, descriptive information and thumbnails of records provided by the members of the *Rede Portuguesa de Arquivos – RPA* (Portuguese Archives Network), facilitating research and retrieval. It is expected that it will be able to integrate other types of services, such as digital preservation;
c) *Ficheiro Nacional de Autoridades Arquivísticas – FNAA* (National Archival Authority File): provides descriptions of creators / holders and their functions, and their relations to external resources held by archives, libraries and museums.

These systems have different architectures, each based on the current descriptive standards created by the International Council of Archives (ICA). The standards govern the collection and processing of information, their management and articulation.

The current archival description standards were created by ICA. The first standard was ISAD(G) (International Council on Archives 2000), which provides general guidance to describe archival material, from the general to the specific, at a multilevel approach. The standard does not provide specific guidance for the description of materials with special characteristics (such as iconography, cartography, architectural, sound or audiovisual), and is more focused on the analog medium, rather on the digital. It refers, in such cases, to specific manuals or standards. The first edition is from 1994, the second from 1999.
It was followed by ISAAR(CPF) (International Council on Archives 2004), intended for the description of creators of archive materials, by ISDF (International Council on Archives 2007), for describing functions, and ISDIAH (International Council on Archives 2008), for the description of institutions with archival holdings. Every standard has the same structure: the information is structured into elements, grouped into different area, according to their typology, referring for a more or less extensive set of ISO standards, to ensure standardisation: e.g. the format of the dates, or the construction of standardised access points.

The fact that the standards have been constructed in a sequential manner, without a previous definition of a basic conceptual model including all the objects to be described, results in their overlapping and inconsistency. Most striking is the existence of two standards for the description of entities - ISAAR(CPF) and ISDIAH. The first one is intended to describe corporate bodies, persons and families as record creators, while the second describes the specific functions of the same type of entities in their condition of archival holders. Hence, both standards have entire areas in common. On the other hand, ISAD(G) provides information on the creation context of the archival records ("Context area"), which is also foreseen by ISAAR(CPF), and ISAAR(CPF) contains information on functions, which have a specific description standard - ISDF.

Standards are supposed to articulate with each other, and there are elements particularly suitable to this purpose - "Reference code" (archival record) and "Authorised form(s) of name" (creators, holders and functions), that are conceived as normalised access points. However, ISAD(G) does not have a relation area, unlike the three other standards, although it has an element to register, in free text, information about related material ("Related units of description"). This is due to the fact it was the first standard, and clearly, because the model of description was not previously defined. In contrast, ISAAR(CPF) and ISDF have two relations areas: one to relate creators or functions, another to relate each of them to archival material and to other resources, or with functions and creators, respectively. As for ISDIAH, it only has a relation area for archival records and its creators.

As the standards were being tested, and it should be noted that the most frequently one used was ISAD(G), ICA itself felt the need to harmonise the four existing standards (International Council on Archives 2012a) or to standardise relationships between the described entities and harmonizing the elements to be used to characterize them (International Council on Archives 2012b). It is relevant to mention the concern in articulate the creators / holders with other types of resources than the archival ones, already expressed in the ISAAR (CPF), in the Identification Area, providing a specific information element for this purpose - "Authorised form of

the name" according to other rules -, in this case rules used by other communities of practice than the archival ones.

In Portugal, the experience of using ICA standards has given rise to different versions of Guidelines to archival description (Direção Geral de Arquivos. Grupo de Trabalho de Normalização da Descrição em Arquivo 2007, 2011). It should be noticed that they provide guidance to describe archival materials, their creators and the choice and construction of access points for corporate bodies, persons, families and geographical entities.

The terminological aspects were a major concern aiming to improve the consistency and the standardisation of the archival descriptions, namely the definition of terms related to the different description areas and their information elements. These documents also provide specific guidelines considering each description levels, since the information to be registered for the same element may be different according to the level of description and it is relevant to "Provide only such information as is appropriate to the level being described", and "To avoid redundancy of information in hierarchically related archival descriptions" (International Council on Archives 2000, 12). A third aspect was the delimitation of the relevant / non relevant information concerning the different elements, - information to consider / information to exclude - wherever possible using concrete examples.

Considering information elements such as "Title" or "Date", the Portuguese Guidelines to archival description recommend the use of attributes (e.g. "formal title", "parallel title", "controlled title", "supplied title", etc., "production date" or "accumulation date") registered in the generic information element "Notes". To standardise the information to register in each element, the guidelines refer, where possible, to specific standards: codes for the representation of names of countries, languages and scripts, date format, bibliographic references, etc.

Considering the number of elements that feature different types of information, the guidelines propose a possible order of its presentation and a syntax to separate them, or at least point to the advantage of defining them. In many cases the advantage of using controlled vocabularies is underscored. One of the most emblematic example is the "Extent and medium" element. Elements such as "History", "Archival history", "Scope and content" are intended to be filled out using narrative text.

"Scope and content" is frequently used to register either information about documentary tradition and documentary typology, or references to marks, seals, inscriptions, signatures, iconography, subjects, entities treated as subjects (corporate bodies, persons, families, geographical entities, etc.), thus demonstrating a rich set of implicit relations between different types of entities. Making these relations explicit, as well as atomising specific information in specific elements, improves the normalisation, the degree of precision of the record and the effectiveness in the recovery of the pertinent information.

The four currently available ICA standards thus point to a vertical approach, considering the area to which they report - only archives - and the objects to to which they relate – archival material, creators, holders and functions -, characterized by some articulation, but by little integration.

Aware of this fact, ICA made available a draft version of a new standard, *Records in Context* (RiC). Intended to be strictly used by archives, it contains a new conceptual model for archival description. It may be described as "*multidimensional description*. Rather than a hierarchy, the description may take the form of a graph or network. Modelling the description as a graph accommodates the single, fonds-based, multilevel description outlined in ISAD(G), but also enables addressing the more expansive understanding of provenance described above. The multidimensional model thus enables the description of the fonds, but also sees the fonds existing in a broader context, in relation to other fonds" (International Council on Archives and Experts Group on Archival Description 2016, 10). This document is on public consultation since early 2017, but the results of the process have not yet been released by ICA. It also lacks the ontology as complementary document, which was initially expected to be available by the end of 2016.

The definition of a new description model is a necessity in the current technological context, characterized by the semantic web, linked data and, consequently, by the need to ensure the interoperability and complementarity of information systems, regardless of the community of practice to which they belong and which holds the resources, be they archives, libraries or museums. For users, it is only necessary to retrieve relevant information, through remote and integrated search in national and international networks and portals (Direção Geral do Livro, dos Arquivos e das Bibliotecas n.d.; Archival Portal Europe Foundation 2016; Europeana Foundation n.d.) using personal computer, but also other devices, such as tablets, or mobile phones.

The ontology CIDOC-CRM (ISO 21127:2014 2014) emerged as a viable alternative, meant to represent information in a wide universe of discourse, since it extends to all areas of human activity, in accordance with the of cultural heritage definition expressed in the Portuguese Cultural Heritage Act (Lei 107/2001 2001).

2 Building the data model

2.1 Methods

Renewing the existing data infrastructure for archival description at DGLAB represents a big challenge, as our core business, archival description, occupies a significant part of the organisation's human resources. At an external level, most of the services we provide, from scientific research to access to legal documents (such as citizenship applications or property documentation), and educational and communication services, rely on our system of archival description. It was clear from the outset that this process should be participative and involve not only decision makers but also end-users, both at internal and external level. Thus, from 2016 to 2017, we held a series of meetings to discuss the new model, based on the needs of each particular community of users.

Work sessions were held with internal collaborators from the two national and one specialised archives in Lisbon and Porto and the 16 regional others on continental Portugal. These focused mainly on the challenges of changing the traditional archival description model and the need to improve the efficiency of services. The compliance of the ICA standards was one of the attendants' main concerns. Several examples from different archive's holdings were chosen by the participants and a great effort was made on mapping its archival description elements to the CIDOC-CRM ontology (from the ISAD (G), ISAAR(CPF), ISDF, ISDIAH and its corresponding metadata schemas EAD, EAC-CPF and EAG (Society of American Archivists and Subcommitee for Encoded Archival Standards 2018; Society of American Archivists and Staatsbibliothek zu Berlin 2014; APEx project 2015). Although the RiC model is not yet stable or formally established as standard, it was included also in these mapping exercises in order to understand its potential compatibility.

Beyond adaptation and transition from a set of guidelines and standards to a new model of representation of knowledge, there was also an interest to understand if the new model could better serve the needs of archival description on topics where the current model seemed either to fail or to present some gaps. Some of these issues related to the establishment of comprehensive relations between information objects, other than the traditional hierarchical ones, or a more accurate representation of time and spatial dimensions of archival records' contextual information. The ability to integrate archival description with management activities such as conservation, document reproduction, reading room orders, and reference assistance was also an important issue to consider. Again this was done based on a practical approach, modelling concrete record sets and testing the possibilities of information representation and the need to develop new features and functions not present on our current information systems.

With external users, the main goal of these meetings was to establish a common ground on requests for improving the current data infrastructure. As it would be expected, these users have, at times, particular needs that are relevant only to small user groups or research projects. The challenge was then to find an equitable solution for all, accommodating particular users without jeopardising the broader approach that should

characterize national archives as a public and inclusive service. The work with external users was carried alongside with internal discussions although there were less working sessions with the former.

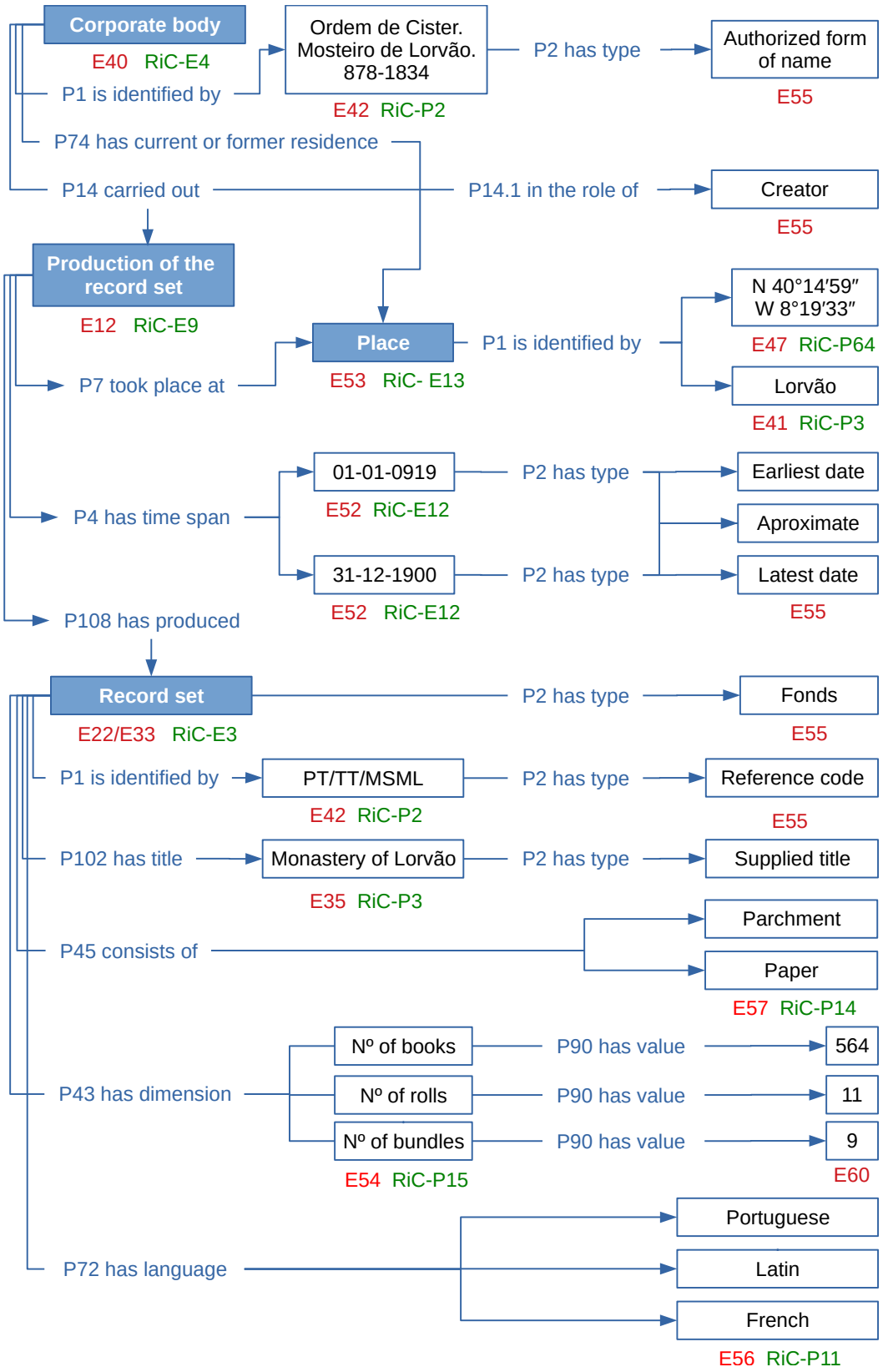

Fig. 1: Fonds representation (selected archival description elements)

2.2 Results

The biggest achievement of the process is probably the involvement of users: even if the technical solutions are to be adjusted on the software development stage, their participation in the data model building process from the outset gave them the possibility of defining concepts from their own experience, which is a clear benefit for an end-user oriented strategy. Other experiences without this approach have proven to be poorly efficient in implementation stages, as users didn't identify the information systems as a response to their needs, but rather as an externally imposed set of rules. The amount of time spent in the discussion process was long but we expect it to be found worthy in the forthcoming phases of the project.

When using CIDOC-CRM classes for representation of archival records and record sets, the first important issue under discussion was the choice of granularity: at what level of detail should knowledge be represented? A record set is undoubtedly a "Man-Made Thing" (E71) and representing it with a broader class is very useful for a wide approach that considers both its physical and conceptual dimensions. However, if subclasses inherit properties from classes on its precedent levels, the reverse can't be applied. Which means that if we want to give information about the languages that can be found in records from an archival fonds, for example, it should be represented as a "Linguistic Object" (E33) as the property "has language (is language of)" can only be applied to this subclass and not to "man-made things" (E33>E73>E89>E28>E71).

Of course this wouldn't be a problem if all the records that compose the fonds were described with the same detail and the group of languages present in the fonds could be retrieved as the result of a search with the appropriate criteria at the level of the record. The issue lies in the fact that that archival description rules and practice imply that this kind of information is often at higher levels on archival hierarchy, prior to the detailed record description. However, one of the reasons that made us decide to use an ontology based model was its flexibility when compared to the traditional hierarchical model: Namely, an instance can belong to more than one class, so the solution is as simple as representing each information object (regardless of its hierarchical positioning in archival description levels) by one or more classes that represent the domain of properties that were found relevant to describe their attributes or elements (Fig. 1)[1]. The granularity of representation is thus determined according to the requirement of use of specific properties, even if this solution sometimes raised some conceptual doubts. The fact that an instance can be represented by different classes also offers the solution for the physical and conceptual duality of information objects in archival description.

Hierarchical part-to-whole relationships are at the core of archival description and one of the main concerns of internal end-users was if it was possible to preserve those relations (and corresponding levels of description) on an object-oriented model such as CIDOC-CRM. Modelling relations from general to specific is quite straightforward as any "Physical Thing" (E18) or "Symbolic Object" (E90), and its subclasses, can be linked to its parts through object property "is composed of (forms part of)" (P46/106). Levels of description can be represented as "Types" (E55) because they clearly are "controlled [terms] used to characterize and classify instances" (ICOM-CIDOC 2015, 27). This is also the understanding of ICA's draft proposal for a new archival description model, that states "Type" (RiC-P23) as one of the "properties of recordset", equivalent to the "level of description, except for the value "item", which equates to the Record entity [RiC-E1]" (International Council on Archives and Experts Group on Archival Description 2016, 26). Using this property to model part-to-whole (and whole-to-part) relationships also opens the possibility to decompose archival records into different

1 In all diagrams, text boxes represent instances. The links between text boxes and corresponding properties are represented by blue characters and arrows; for graphic clarity purposes, relations are expressed only in their forward name, although, in most cases, could be expressed in their reverse as well. Also for the sake of clarity, class labels are omitted. The matching CIDOC-CRM classes for instances are written in red characters below text boxes. When more than one occurs, they are separated by / . The equivalent RiC codes (entities and properties as defined in the draft document) are written in green characters. The diagrams represent a version of mapping and modelling examples simplified according to the goals of the current text.

components beyond the traditional lowest archival description level. This is very useful in representing components such as miniatures on an illuminated codex such as the *Apocalypse of Lorvão* (Fig. 2), or digital objects attached to an e-mail message, for example.

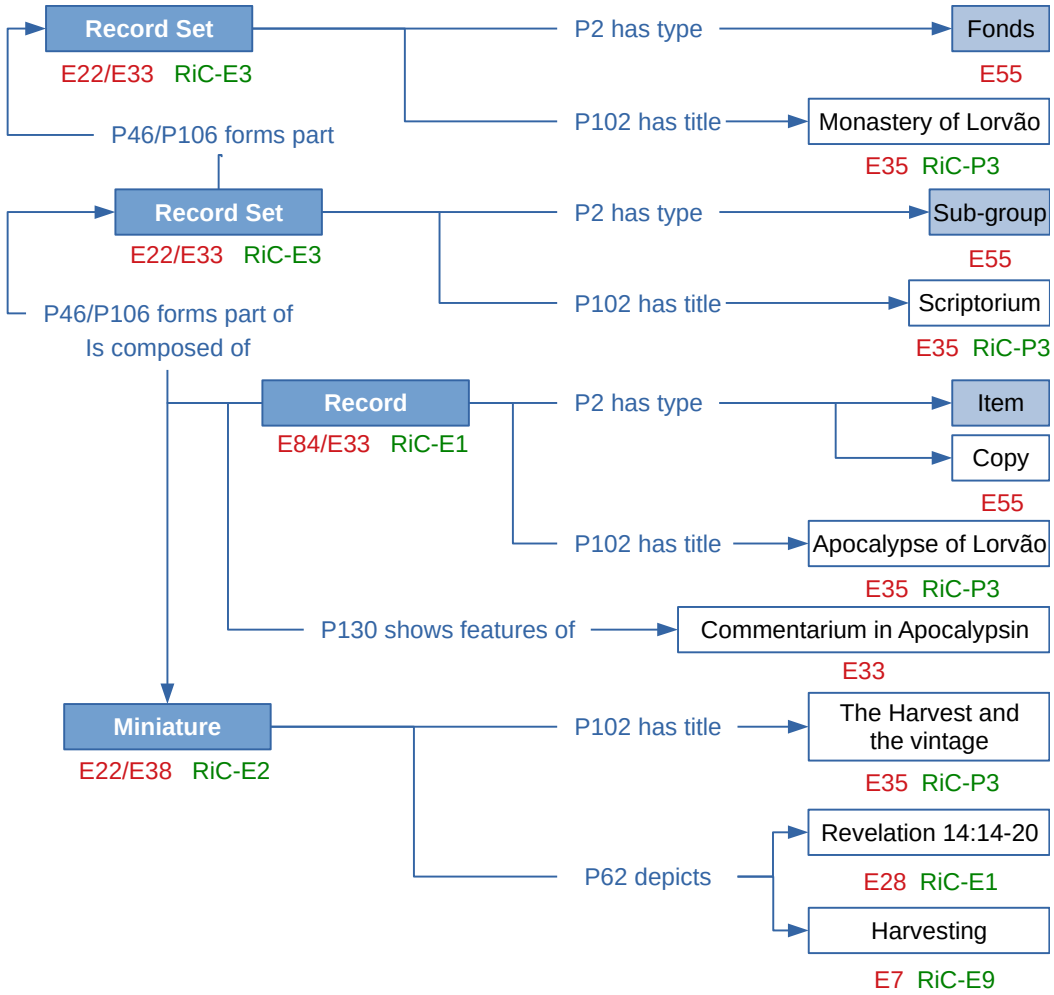

Fig. 2: Modelling part-to-whole relationships and corresponding archival description levels

The definition of "Activity" (E7) is conformable with the concept of *function* as defined on ISDF (International Council on Archives 2007). Some of its subclasses can represent functions of corporate bodies associated with the creation and maintenance of archives, like "Transfer of Custody" (E10) or "Acquisition" (E8) for example. However, those subclasses are not enough to represent the complexity of functions associated with archival records which, in many cases, are directly related to administration and business processes. Functions are better described with terms from thesauri and controlled vocabularies (National Archives of Australia 2016; Direção Geral do Livro, dos Arquivos e das Bibliotecas 2018, as examples in English and Portuguese). Therefore it seemed more accurate to use the class "Type" (E55) for most of archival functions, linked to their corresponding activities by the property "had general purpose (was purpose of)" (P21). Some functions are also inferred by the type of role actors have in events and activities and can thus be represented by the typed properties like "some actor (E39) participated (P11) in some event (E5) in the role of (P.11.1) some type (E55)" or "performed (P14) some activity (E7) in the role of (P14.1) some type (E55)". Instances of these types should be "listed in (P71) some authority document (E32)".

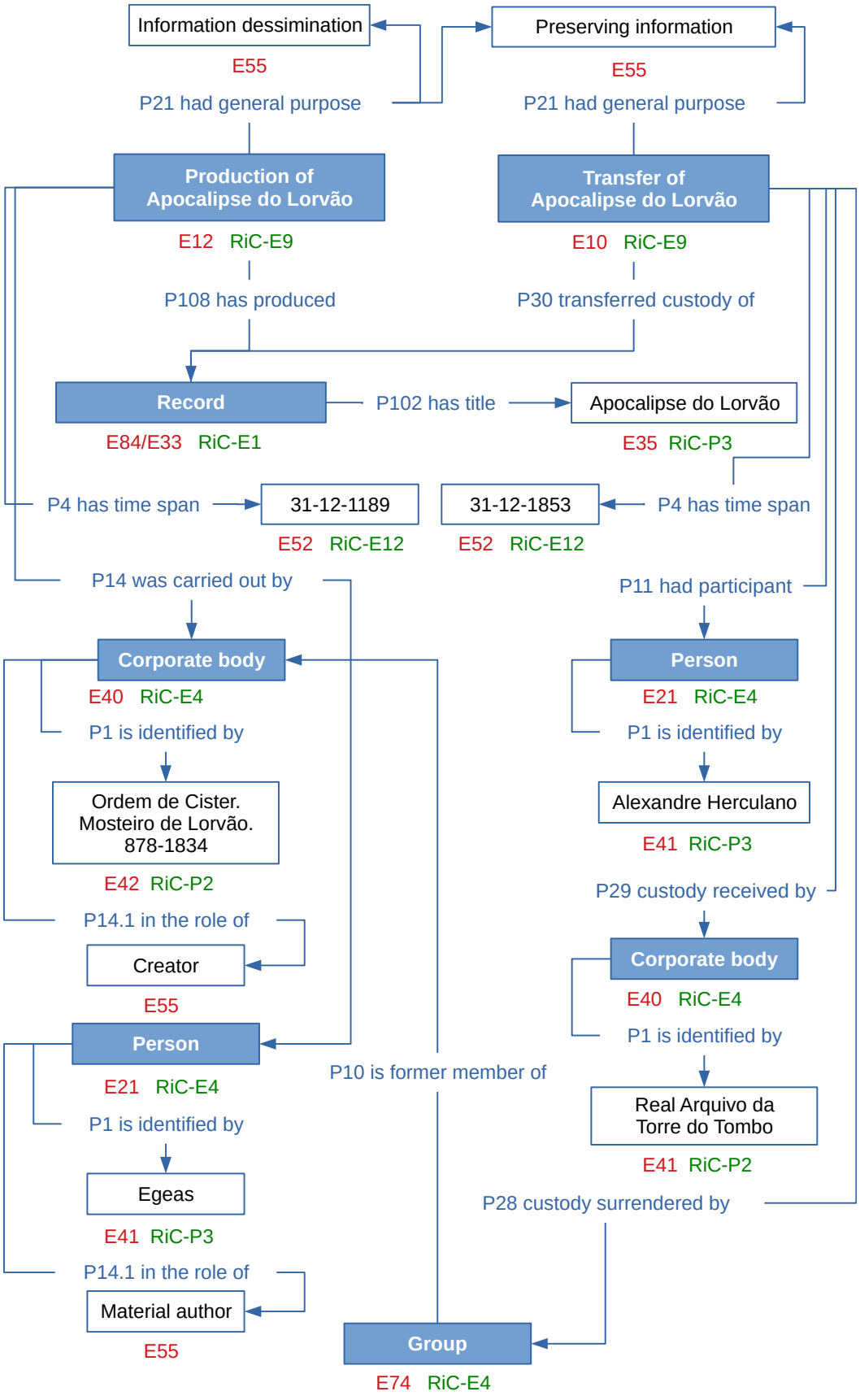

Fig. 3:Representation of activities and archival functions

The concept of *mandate*, as the source of authority to perform functions (International Council on Archives 2004, 19–20, 2007, 10, 2008, 25–26), has been proposed to be modelled using class "Design or Procedure" (E29) (Hennicke 2013) but it also can be modelled by other immaterial items that fall within the class "Conceptual Object" (E28) or another of its subclasses. The chosen representation should be the one more suitable for each *mandate*, a concept that covers a significant range of conceptual items, from contemporary legislation to belief systems or socially approved practices. Any of those items can be the reason, or mandate, for carrying out the some "Activity" (E7), described by the property "was motivated by (P17) some conceptual object (E28)".

A bigger challenge was to model archival description elements that represent corporate bodies, persons and families (International Council on Archives 2004, 2008). All the information required by these standards can be represented by the class "Actor" (E39), and its subclasses, through their participation in several events (E5) or activities (E7). The question that arose from practical applications of this representation was the complexity of the consequent graph.

The relations between members of the same family, apart from the direct ones that are established through "Birth" (E67), can be tangled: to retrieve the information that two persons (E21) are cousins, for instance, there is no direct path to link them in that condition. Their family relation can only be inferred through the participation in a series of events: their own births, but also that of their parents' (that establish that at least two of the four are siblings), marriages or other family associations. Even if we choose to model these relations by the membership within a family group, without their participation in events, using typed properties such as "kind of member" (P107.1) (ICOM-CIDOC 2015, 84), this again is a complex solution: they can have different types of membership that need to be distinguished according to the member(s) of the family group(s) they are related to (if they are cousins, simultaneously are nephews of one another's parents, and so forth). Therefore, the representation, or inference, of the relations between members may imply the multiplication of subgroups within a family group. This complex and tangled representations are found as well in corporate bodies or other gatherings or organizations of "Actors" (E39). We understand that this is a well known issue (Ore 2014) and look forward to the development of CRMsoc (Doerr 2018) in order to better address this question.

The graph representation of other archival description elements has the same complexity in cases where information is stored in text strings which aggregate multiple instances of different natures. This is true mainly for "Administrative / Biographical history", "Archival history", "Immediate source of acquisition or transfer" and "Scope and Content" (International Council on Archives 2000, 18–24). This approach is kept in the draft proposal for a new conceptual model for archival description that still recommends the use of text strings or notes to represent, for example, a "summary of scope (such as, time periods, geography) and content (such as subject matter, administrative processes) of the Record" (International Council on Archives and Experts Group on Archival Description 2016, 23).

The atomisation of these text strings into their different elements and the representation of the same information by semantic statements implies a significant shift in archival description practices. Context, content, and structure information about archival records are traditionally represented by descriptive texts, as well as conditions of access and event information about the production of the description itself. These text strings are bound by ICA standards, and subsidiary rules and regulations (in Portugal, see Direção Geral de Arquivos. Grupo de Trabalho de Normalização da Descrição em Arquivo 2011) thereby guaranteeing a certain degree of normalisation of their structure and terminology. Even so, it is not easy to avoid data redundancy and/or guarantee data integrity, and their negative impact on data retrieval.

Modelling these elements with CIDOC-CRM implies the representation of several events that were not stated in previous descriptions. All information about dates of production and accumulation of record sets, or establishment, foundation and dissolution of corporate bodies, for example, is related to events and activities that were only implicit on the text strings. Representing these events contributes for the semantic enrichment of

archival description and data integrity but, again, is an important change on description methods. The work with internal end-users showed that probably this change will be the largest obstacle on implementation stages. However, this is a problem on procedures and practices that we expect to overcome with training and a progressive identification of users with a more efficient information system.

In all modelling and mapping exercises we identified a recurrent need to use the class Type (E55) to represent many archival description elements. Although this is not surprising considering the definition and modelling principles of "types" (ICOM-CIDOC 2015, xvi–xviii), it underscored the obligation to choose, adapt and, in some cases, translate existing thesauri and controlled vocabularies for archival description. These domain specific terminologies will be represented as instances of Type (E55), linked by the property "has broader term (has narrower term)" (P127) when terms hierarchies are applicable, and "listed in" (P71) "Authority documents" (E32).

3 Discussion and further developments

Building the data model was a challenging process that helped us clarify concepts and establish the bases for a more efficient information system. The development of a CIDOC-CRM core data model for DGLAB was undertaken on a conceptual level and, at times, the absence of a concrete implementation solution hindered the communication with with end-users, both internal and external.

In order to address this issue, we initiated a cooperation with Institute for Systems and Computer Engineering, Technology and Science (INESC-TEC), Informatics Engineering Department, Engineering Faculty of the University of Porto (FEUP) on a SR&TD Project Data Science and Artificial Intelligence for Public Administration, funded by *Fundação para a Ciência e a Tecnologia, I.P.* (FCT), the Portuguese Foundation for Science and Technology. The EPISA project (*Entity and Property Inference for Semantic Archives*) aims, amongst other goals, to develop a prototype for an open-source knowledge graph platform adopting our data model for archival description. Due to the complexity of the paradigm shift from representing archival information on a linked data model, this project is also devoted to finding ways to guarantee the effective migration of contents stored according to ICA standards to an ontology based model, requiring both the use of existing cross-walks and the inference of the new relations with semi-automated methods.

Replacing the current data infrastructure also raises the important challenge of migrating existing descriptions. The results of the mapping exercises will be the starting point to define a cross-walk for direct mappings, using tools already tested between different metadata schemes, including EAD (Bountouri and Gergatsoulis 2011; Gaitanou, Bountouri, and Gergatsoulis 2012). Beyond direct mappings, as ontology modelling enables the inference of ontological relations (Hitzler, Krötzsch, and Rudolph 2010), we intend as well to determine if it is possible to infer new assertions through analogical reasoning based on our data model. The overall goal is to focus on the relevant classes identified in the ontology, for which information is scattered in several elements of the original descriptions, and extract meaningful values for the relevant properties previously identified in the ontology. As it has been said, most of metadata records includes information stored in long text strings so an effort will be done to use natural language processing tools (such as syntactic analysis, named entities recognition and relation extraction) to obtain semantic representations for selected parts of the metadata records, interpreting them in the context of cultural heritage using the team's expertise in other knowledge fields (Mendes et al. 2014).

The starting point of ICON Project was the renewal of the existing data infrastructure for archival description, document conservation and preservation, access to information and its workflows. From the early stages of our work we focused on interoperability and content integration and the development of the project made us believe that it can be one of its key features: we intend to create a more flexible data model that can both interoperate with other information systems and accommodate information of cultural resources other than

archival documents. Representing cultural objects through a semantic approach seems to be the answer to the efficiency of our response to internal and external users in a world where the ubiquitous connectivity of everything emerges as a reference value (Curvelo et al. 2014). The conceptual work we undertook so far showed that most of our problems and found solutions for data and knowledge representation are not exclusive of the archival description universe: therefore, we hope the ICON Project can effectively contribute to this trend with the development of concepts and tools than can be transversally applied in different cultural heritage management institutions.

4 References

APEx project. 2015. *EAG 2012: Encoded Archival Guide. Work Packages 2 & 4*. [Den Haag]: APEx project. http://www.apex-project.eu/index.php/en/outcomes/standards/eag-2012.

Archival Portal Europe Foundation. 2016. 'Archives Portal Europe'. 2016. https://www.archivesportaleurope.net/home.

Bountouri, Lina, and Manolis Gergatsoulis. 2011. 'The Semantic Mapping of Archival Metadata to the CIDOC CRM Ontology'. *Journal of Archival Organization* 9 (3–4): 174–207. https://doi.org/10.1080/15332748.2011.650124.

Curvelo, Paula, ângela Guimarães Pereira, Philip Boucher, Melina Breitegger, Ghezzi Alessia, Caroline Rizza, Mariachiara Tallacchini, and Lucia Vesnić Alujević. 2014. *The Constitution of the Hybrid World: How ICT's Are Transforming Our Received Notions of Humanness*. https://doi.org/10.2788/58678.

Decreto-Lei 102/2012. 2012. *Diário da República*. Vol. 92/2012. http://data.dre.pt/eli/dec-lei/103/2012/05/16/p/dre/pt/html.

Direção Geral de Arquivos. Grupo de Trabalho de Normalização da Descrição em Arquivo. 2007. *Orientações para a Descrição Arquivística: 2ª versão*. Lisboa: Direção-Geral de Arquivos.

———. 2011. *Orientações para a Descrição Arquivística: 3ª versão*. Lisboa: Direção-Geral de Arquivos.

Direção Geral do Livro, dos Arquivos e das Bibliotecas. 2018. 'CLAV – Plataforma para a Classificação e Avaliação da Informação Arquivística'. 2018. http://clav.dglab.gov.pt/.

———. n.d. 'Portal Português de Arquivos'. Accessed 7 August 2018. https://portal.arquivos.pt/.

Doerr, Martin. 2018. 'Issue 358: CRMsoc and Scope of CRM Modules | CIDOC CRM'. *Crm-Sig*. http://www.cidoc-crm.org/Issue/ID-358-crmsoc-and-scope-of-crm-modules.

Europeana Foundation. n.d. 'Europeana Collections'. Europeana Collections. Accessed 7 August 2018. https://www.europeana.eu/portal/?locale=en.

Gaitanou, Panorea, Lina Bountouri, and Manolis Gergatsoulis. 2012. 'Automatic Generation of Crosswalks through CIDOC CRM'. In *Metadata and Semantics Research*, 264–75. Communications in Computer and Information Science. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-35233-1_26.

Hennicke, Steffen. 2013. 'Representation of Archival User Needs Using CIDOC CRM'. In *Workshop Practical Experiences with CIDOC CRM and Its Extensions (CREMEX 2013)*, 48–60. Valetta. https://pdfs.semanticscholar.org/ba2e/79d67c0b9d13bb7cc8fc6172ee7eb39a18de.pdf.

Hitzler, Pascal, Markus Krötzsch, and Sebastian Rudolph. 2010. *Foundations of Semantic Web Technologies*. Chapman & Hall/CRC Textbooks in Computing. Boca Raton: CRC Press.

ICOM-CIDOC. 2015. *Definition of the CIDOC Conceptual Reference Model*. Edited by Patrick Le Boeuf, Martin Doerr, Christian Emil Ore, and Stephen Stead. Version 6.2. ICOM/CIDOC Documentation Standards Group. http://www.cidoc-crm.org/sites/default/files/cidoc_crm_version_6.2.pdf.

International Council on Archives. 2000. *ISAD(G): General International Standard Archival Description: Adopted by the Committee on Descriptive Standards Stockholm, Sweden, 19-22 September 1999*. 2nd ed. Otawa: International Council on Archives.

———. , ed. 2004. *ISAAR(CPF): International Standard Archival Authority Record for Corporate Bodies, Persons and Families*. 2nd ed. Standards. Canberra Australia: International Council on Archives.

———. 2007. *ISDF: International Standard for Describing Functions Developed by the Committee on Best Practices and Standards, Dresden, Germany, 2-4 May 2007*. 1st ed. Paris: International Council on Archives. https://www.ica.org/en/isdf-international-standard-describing-functions.

———. 2008. *ISDIAH: International Standard for Describing Institutions with Archival Holdings: Developed by the Committee on Best Practices and Standards London, United Kingdom, 10-11 March 2008*. Paris: International Council on Archives. https://www.ica.org/en/public-resources/standards.

———. 2012a. *Progress Report for Revising and Harmonising ICA Descriptive Standards*. International Council on Archives. https://www.ica.org/en/cbps-progress-report-revising-and-harmonising-ica-descriptive-standards.

———. 2012b. *Relationships in Archival Descriptive Systems*. Paris: International Council on Archives. https://www.ica.org/en/cbps-relationship-archival-descriptive-systems.

International Council on Archives, and Experts Group on Archival Description. 2016. *Records in Context: A Conceptual Model for Archival Description*. Consultation Draft v0.1. International Council on Archives. https://www.ica.org/en/egad-ric-conceptual-model.

ISO 21127:2014. 2014. *Information and Documentation. A Reference Ontology for the Interchange of Cultural Heritage Information*. 2nd ed. Information and Documentation. Genebra: International Standard Organization. https://www.iso.org/standard/57832.html.

Lei 107/2001. 2001. *Diário da República*. Vol. 209/2001. http://data.dre.pt/eli/lei/107/2001/09/08/p/dre/pt/html.

Mendes, David, Irene Pimenta Rodrigues, Carlos Rodriguez-Solano, and Carlos Fernandes Baeta. 2014. 'Enrichment/Population of Customized CPR (Computer-Based Patient Record) Ontology from Free-Text Reports for CSI (Computer Semantic Interoperability)': *Journal of Information Technology Research* 7 (1): 1–11. https://doi.org/10.4018/jitr.2014010101.

National Archives of Australia. 2016. 'Australian Governments' Interactive Functions Thesaurus (AGIFT)'. 2016. https://data.naa.gov.au/def/agift.html.

Ore, Christian Emil. 2014. 'Ontologies for Prosopography: CIDOC-CRM'. Workshop presented at the Digital Humanities 2014 conference, Lausanne. http://edd.uio.no/artiklar/DH2014/C-E_Ore_prosopography.pdf.

Portaria n.º 192/2012. 2012. *Diário da República*. Vol. 117/2012. http://data.dre.pt/eli/port/192/2012/06/19/p/dre/pt/html.

Society of American Archivists, and Staatsbibliothek zu Berlin. 2014. *EAC-CPF: Encoded Archival Context - Corporate Bodies, Persons, Families: Schemata and Tag Library*. Berlin: Staatsbibliothek zu Berlin. http://eac.staatsbibliothek-berlin.de/schemata-and-tag-library/.

Society of American Archivists, and Subcommitee for Encoded Archival Standards. 2018. *EAD: Encoded Archival Description. Version EAD3 1.1.0*. Chicago: Society of American Archivists. http://www.loc.gov/ead/eadschema.html.