

Importance of data standardization and the NEDA-CM implementation in a repository of archival entities: PARES

Ana López Cuadrado

Culture and Sport Ministry

National History Archive

Address: Serrano 115, (Spain), 28006

anamaria.lopez@cultura.gob.es

Abstract

Due to changes produced in our society and above all in our researchers, and the way in which they search in our documentary heritage as a primary source, standardization has been considered by the archival community as the key for improvement of data processing. Internet, the new technologies and social media are great opportunities if we know how to take advantage of it. This is the issue in which the Spanish National Archives have worked in the last years. According to the international and national archival standards, we have developed a tool, PARES, which allows archivists to describe data in standardized way and users to search these data in an easy way. Conclusions of this work have been clear: if we need researchers find data without an archivist beside him, we need not only technology and archival knowledge but also standardization of all our processes, that is, we must describe taking into account the archival standards.

Keywords: archival knowledge, archival standards, digital technology, semantic web, Record in Context

This article summarizes the presentation made at workshop “Trust and Understanding: the value of metadata in a digitally joined-up world” organized by DARIAH Working Group on Sustainable Publishing of Metadata, held in Brussels, on May, 14-15, 2018.

Information to access is the freedom or ability to identify, obtain and make use of data or information effectively for human users and, in this way, accessing and further processing large and unwieldy amounts of data and information.

In addition, information to access covers many issues including copyright, open source, privacy, and security.

But what is the status of the issue regarding archival information to access? What happened in the last ten years to reach the current situation?

Let's review:

We can distinguish 4 issues which explain why we have decided to start thinking in standardization in our data bases:

- First: Changes in the profile of our users and the way in which they make the information to access. Our stats show clearly how each year we had less users in our "search room" while we had a lot of virtual users which demanded us information, which many times they could find on line, but the self-search was difficult. We started to think we had built our Portal with an archival vocabulary, thinking in users got used to find records with the help of archivists. Problem: archivists aren't in their side when they make search online. Result: they don't find data so easily and they send a lot of emails asking for the same information many times. It was the moment to improve our Portal if we wanted to reduce questions. Consequently, we could work in other issues to offer more new data, instead of working only answering all the same questions we received.
- Changes due to the emergence of national and international archival rules. A lot of international rules have appeared in the last 10 years. For this reason, archivists have been to adapt their description way to all of them.
- Multilingualism. In Spain we have one official language and three co-officials, and on the other hand we know we receive questions from different parts of the world. One of the main problems was the information to access because of the language. For this reason, we started to show the same information, not only in the official and co-official languages of Spain, but also in English and French. As a result, we could achieve more users and get a Portal more intuitive and accessible for all.
- IT developments (Semantic Web, Digital Humanities, Big Data). The IT changes should be an option to improve not a problem, but in many cases, it happens the second. Archivists should take advantage of this situation and get more automatically processes and better tools with the help of our IT developers. The obligation for the archivists is to know these improvements and to study which could be interesting for us, depending on our necessities.

In this IT environment, the archival institutions were delayed in many cases, if we look at other cultural institutions. The current situation in archival information to access is clear but “sad”. In many archival institutions, we can find many or all of these points:

- Single or multiple relational databases, which are free access to users, but many data bases what means problems to archivists to give a quick answer.
- New and different archival rules (archivists have problem to use all of them and to distinguish when they should use each one).
- Highly qualified professional staff with specialized knowledge but with problems to move, apply or understand the new technologies.
- Data: information very interesting, useful to the citizens, not only to the researchers, but without defining or standardizing; noise and lack of organization in our tools.

For this reason, we concluded thinking it was necessary archival standardization to achieve a clear improvement in information to access; but the next questions were: **what is the standardization of archival description? Which are the benefits? And the most important, how should we carry out this archival description in standardization way?**

What it is?

Standardization archival description can be defined as the description of archival entities in an integral way, taking into account the archival standards, allowing the exchange of data between different archival systems and thinking “a priori” how facilitating to the users the search of that information. I mean, describing thinking in finding, not only in arranging (in the first sight could be seemed the same but you will obtain different results depending on your way describe the records created by one institution, person or family.

Which are the benefits?

The benefits are simple but very important for the information to access. If we use standardized archival entities, we will achieve:

- Making easier the understanding of archival records, showing their context, content and structure.
- Making possible the integration of archival descriptions from different locations in a unified and shared information system
- Improving archival information to access, data search and data use and reuse

- Information exchange between institutions with similar content, which will reduce time and workload to be more effective. In this way it will be optimized the results obtained by others in one side and in other side shared our results.

How should we standardize?

It can be distinguished and use three blocks of archival standards.

- **Content Standards**

International Archival Council (ICA) has developed four international standards to describe archival entities:

- ISAD (G)¹ (1^a ed. 1994; 2^a ed. 2000) to describe records in a multilevel way
- ISAAR (CPF)² (1^a ed. 1994; 2^a ed. 2004) to describe agents
- ISDF³ (2007) to describe functions
- ISDIAH⁴ (2008) to describe archival holdings as agents

- **Communication Standards**

They appear to facilitate data interchange

- EAD⁵ : to describe records (supported by SAA)
- EAC-CPF⁶ : to describe agents (supported by SAA)
- EAG 2012⁷ : to describe archival holdings as agents (supported by APEF)

¹ ISAD (G): International Standard Archival Description, Second Version (Sub-Committee on Describe Standards). See more

<https://www.ica.org/en/isadg-general-international-standard-archival-description-second-edition>

² ISAAR (CPF): International Standard Archival Authority Record for Corporate Bodies, Persons and Families, 2nd Edition (CBPS - Sub-Committee on Descriptive Standards). See more:

<https://www.ica.org/en/isaar-cpf-international-standard-archival-authority-record-corporate-bodies-persons-and-families-2nd>

³ ISDIF: International Standard for Describing Functions (CBPS - Sub-Committee on Descriptive Standards). See more: <https://www.ica.org/en/isdf-international-standard-describing-functions>

⁴ ISDIAH: International Standard for Describing Institution with Archival Holdings (CBPS - Sub-Committee on Descriptive Standards). See more:

<https://www.ica.org/en/isdiah-international-standard-describing-institutions-archival-holdings>

⁵ Encoded Archival Description. See more:

<https://www2.archivists.org/groups/technical-subcommittee-on-encoded-archival-standards-ts-eas>

⁶ Encoded Archival Context. See more:

<https://www2.archivists.org/groups/technical-subcommittee-on-encoded-archival-standards-ts-eas>

⁷ Encoded Archival Guide: See more: <http://wiki.archivesportaleurope.net/index.php/EAG2012>

- **High-level Conceptual Standards**

- NEDA Conceptual Model⁸ (Spanish Archival Conceptual Model)
- RIC: Record in Context (ICA Conceptual Model⁹)

The most important point is that all these standards are matched, I mean, we must use both in our process to get a complete result. But in this article, I would like to focus on the last type, the high-level conceptual models. Due to technological changes, new systems to show or represent the information and the change in the archival description taking into account multiple agents, it was necessary to create Conceptual Models of Archival Description, which allow a new way to understand the archival science. An archival conceptual model wants to represent all the elements the archivists need to describe, how those elements are interconnected or interacted each other, and data we need to describe them. Both projects, Records in Context and NEDA-CM grew with those premises and have common elements. Record in Context used, the Spanish conceptual model to develop its philosophy. For this reason, when we say “we use NEDA Conceptual model to describe our records¹⁰”, we mean we use the philosophy of the RIC, or an integral philosophy to describe our documentary, but we can’t forget we use content and communication standards too. The use of all of those plus the new technologies and the standardization policies allow us achieving the principal goal of this article, the improvement of the information to access in the archival environment¹¹.

To sum up, it can be concluded standardization of archival description deals with identification and description of different archival entities, applying the international standards defined previously, taking into account the multilevel description showing by ICA standards and adding the new archival conceptual model (multi-entity description).

Now that we know what is an archival conceptual model and before continuing with the explanation, we should need to understand other concept base in the new archival science. It’s

⁸ NEDA-CM: Modelo Conceptual de Descripción Archivística. See more:

<https://www.mecd.gob.es/cultura/areas/archivos/mc/cneda/documentacion/normas/neda-mc.html>

⁹ Record In Context-Conceptual Model (Expert Group on Archival Description). See more:

<https://www.ica.org/en/egad-ric-conceptual-model>

¹⁰ The objective of CNEDA is to provide assessment in the standardisation of the Archival Description, developing and updating the Spanish Standards for Archival Description (hereinafter, NEDA), which are defined as an applicable regulatory tool aimed at the continuous improvement of access to the archival records. Between 2007 and 2017, CNEDA drafted the Conceptual Model of Archival Description (hereinafter, NEDA-MC), together with this technical glossary (hereinafter, NEDA-Voc). See more: https://www.mecd.gob.es/dam/jcr:c79496cf-1a09-4bb0-a977-f350b032d468/NEDA-Voc_eng.pdf

¹¹ Requejo Zalama, Javier RiC-CM y MC-NEDA: ¿Nueva Descripción Archivística? TRIA Nº 21. 2017 I.S.S.N. 1134-1602 - Págs. 79-95

necessary defining what is “multi-entity description” and why this term is so important understanding well to getting better descriptions in our archival system.

According to NEDA Conceptual Model, entity is any real or abstract object which exists, has existed or may come to exist. Entities identified in NEDA-MC include Records, Agents, Functions, Mandates, Places and Concepts, Objects or Events.

When you make an archival multi-entity description, you should take into account all the issues which appear in the document, and how is the relationship of those with that. I mean, the place in which a record is made, the person who signs it, people or institution which appear in the text, the historical moment, why somebody made the document, or who somebody ask for writing it, who writes it, who asks for, why?, where?, even why is this record in the archive or which process of one institution is showed in this document?. Many questions you have to resolve when you describe if you want to understand perfectly the record. The context which goes around the document which explains or completes the real meaning of the record.

All these data can be showed in our database through entities and relationships. I mean, your document is made in a place and sign, for example, for a king, but in this place and this king surely have made more records... If you make a relation between the description in your database of your document and the description of the king or the description of the place, you will achieve users can navigate in your database finding more information related to their records, and maybe discovering data which they unknown. This is the Record in Context. The study of the document through the “entities” which are interrelated and their representation through a computer program.

Only one more issue. The multi-entity description isn't incompatible with the multi-level description, on the other hand both are completed. You can continue showing the multilevel description if you indicate in your description of the record who is the creator of the record, or if you made a relationship between the document and the series in which this record is, or the fond (group of records).... Even so, when you describe, you should make a relationship for example between the institution which creates the document, the functions of that institution, more documents that are created by the institution, the series that are formed by all the records which are generated by that institution when it makes that function.... In Summary, the multi-entity description includes the multi-level ones.

It's also important to explain that describing with access points is totally contrary to the multi-entity concept. When you describe, you don't add subjects or access points in general in the record. When you describe from a multi-entity concept, you must go further. First you

must think about the entities which are related with the document, then you will must guess why those are related, and finally you should show this relationship in your tool.

If we describe taking into account the multi-entity concept, we will identify, study and deal with all the archival entities, leaving records in the middle of the circle, I mean again, Records in Context, and this way of understanding the records allows us getting completely the meaning of our heritage, and show it in a comprehensible way to the users.

In order to carry out this task, Spanish Archival States have developed PARES, our records description system, which allows to describe all the entities defined by NEDA Conceptual Model (dashboard) and an efficient and effective information to access in a fast and intuitive way.

PARES¹² has become the tool through which the Spanish conceptual model and their archival description philosophy has been launched. PARES was conceived as a relational database, but it has been improved thanks to applying the Spanish Conceptual Model. It was born to describe records, but it has become, thanks to the effort of many people, in a tool that describes all kind of entities, related between them and linked to external data. In addition, it has been taken the leap capturing data automatically through APIs from other cultural institutions and the different technologies born around the semantic web.

When archivists start to describe a record, they study it (as a primary source) taking into account its context. All those entities which are represented in the record are relevant and must be related to each other and to records. In that way, our records system grows with ton of related entities and links between them. In that way, we will be able to create a real navigation effect. This interconnection will be benefited exponentially if we not only reuse the information but also become content providers, this is: we convert our system into linked open data.

PARES demonstrates the Records in Context or the multi-entity description can be real. This article only wants to show the archival standards aren't pure theory, but without forgetting the way in which our colleagues have described the previous years. The classic archival description methodology based on catalogue cards, it is really useful if we take advantage of all the technical and normative tools at our reach, achieving an improvement in the information to access. I mean, how people totally distant to the archival vocabulary and without an

¹² <https://pares.mecd.gob.es/inicio.html>

intermediary archivist, could search information quickly, without problems, and with a tool very close to google search, only using the multi-entity concept and the new technologies.

In the Spanish case, we have much more issues to improve and solve, but we consider this is a path. We must to continue step by step, but always, thinking in users (information to access) and in an integral way: the record and its context. But nowadays PARES works ok, and citizens and researches are happy with that.

So, archivists, please!!! Let's start to work hand in hand with our IT colleagues. They will understand very well our new archival standards, in many cases better than us. In this way we will achieve more people know our documentary heritage, and from this view we will be able to show its value, and it will be easier preserve it for us.