

**Trust and Understanding :
the value of metadata
in a digitally joined-up world**

Published with the financial support of



And in collaboration with



ARCHIVES ET BIBLIOTHÈQUES DE BELGIQUE
ARCHIEF- EN BIBLIOTHEEKWEZEN IN BELGIË



NUMÉRO SPÉCIAL 106 EXTRANUMMER

**Trust and Understanding :
the value of metadata
in a digitally joined-up world**

Rolande DEPOORTERE, Tom GHELDOLF,
Dorien STYVEN, Johan VAN DER EYCKEN (eds.)

Brussel/Bruxelles
2019

ARCHIVES ET BIBLIOTHÈQUES DE BELGIQUE
ARCHIEF- EN BIBLIOTHEEKWEZEN IN BELGIË



First published in Belgium 2017
p/o Archives générales du Royaume/Algemeen Rijksarchief
Rue de Ruisbroeck / Ruisbroekstraat 10
B 1000 Bruxelles/Brussel

A catalogue record of this book is available from the Royal Library of Belgium

Dépôt légal/Wettelijk Depot D/2019/1080/2
ISSN : 0775-0722
© ARCHIEF- EN BIBLIOTHEEKWEZEN IN BELGIË
ARCHIVES ET BIBLIOTHÈQUES DE BELGIQUE
Bibliothèque Royale de Belgique/Koninklijke Bibliotheek van België
Boulevard de l'Empereur/Keizerslaan 4
B 1000 Brussel/Bruxelles

All rights reserved. No part of this publication may be reproduced, stored in a revival system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of ABB

IMPORTANCE OF DATA STANDARDIZATION AND NEDA-CM IMPLEMENTATION IN AN ARCHIVAL REPOSITORY: PARES

Ana LÓPEZ CUADRADO

Introduction

Access to information is the freedom or ability to identify, obtain and make use of data or information effectively for human users and, in this way, to access and further process large and unwieldy amounts of data and information. In addition, *access to information* covers many issues including copyright, open source, privacy, and security. Through this article we wish to go into the state of affairs regarding the issue of access to archival information. We will discuss the developments which took place in Spain for the past ten years and present an overview of the current situation.

In 2007, the Spanish State archives took the decision to standardize its databases. This decision was motivated by four issues: The profiles of our users had changed as well as the way in which they accessed information. Our statistics clearly showed a decreased use of our *search room*, while many online users requested information that they easily could have found online. Carrying out the research themselves appeared to pose some difficulties. This made us realize that we had worded the texts of our Portal in rather technical archival vocabulary, while thinking that users had gotten used to find records with the help of archivists. One of the problems of conducting online research is the absence of archivists. They cannot assist users for each step they have to take during online research. As a result users did not find data so easily. We received a lot of emails asking for the same information over and over again. This encouraged us to improve our Portal in order to reduce written research requests. As a result, we were able to work on other issues and to provide new data instead of only answering the same questions.

Changes due to the emergence of national and international archival rules - a lot of those were introduced during the last ten years - forced archivists to adapt their methods and the way to describe archives accordingly.

Spain has one official language and three co-official languages. The archives also receive questions from different parts of the world. Consequently, the language barrier became a key issue of *access to information*. For this reason, we decided to provide the same information, not only in the official and co-official languages of

Spain, but also in English and French. As a result, we were able to increase the number of users and to make the Portal more intuitive and accessible for all.

Advances in information technology (Semantic Web, Digital Humanities, Big Data) should be considered as means for progress, not as a problem. However, in many instances, the advancement of IT raises a considerable amount of issues. Archivists should take advantage of this situation in order to further develop automation processes and improve tools with the help of IT developers. Archivists should be aware of these developments and study how they could be of interest for the archival community and how they could meet the needs of the community.

Compared to other cultural institutions, archives have fallen behind with regard to information technology. In the current situation, access to archival information is far less than optimal. Archive institutions are often confronted with the same problems. They manage and store different single or multiple relational databases, which users can access freely. But the amount of databases to be searched makes it difficult for archivists to reply quickly to research requests. Archivists are constantly confronted with new and different archival rules. This makes it difficult to apply them all or to make a distinction which one should be applied and which one not. Highly qualified and professional staff with specialized knowledge face problems in moving, applying and understanding new technologies. Archives dispose of very interesting and useful information for citizens, not only for researchers. However, without definition or standardization tools lack precision and organization.

The need for standardization

These four reasons made us realize that it was necessary to introduce archival standards in order to significantly improve access to information. *The standardization of archival descriptions can be defined as the description of archival entities in an integral manner, taking into account archival standards and allowing the exchange of data between different archival systems.* But concomitant questions already arose: in which way are archive descriptions best standardized? Which are the benefits? And, most importantly, how to apply archival description in a standardized manner? The user must also be taken into account. How can we optimize and facilitate the search for information? The manner in which descriptions are thought out - not only their arrangement - influences search results. Descriptions who may seem similar at first sight lead to other results.

The benefits of standardization are obvious. Standardization is very important to be able to consult information, as it will:

- be easier to understand archival records, show their context, content and structure.

- enable integration of archival descriptions from different origins in a unified and shared information system
- improve access to archival information, data research and data use and reuse
- achieve information exchange between institutions about similar contents, which will reduce time expenditure and workload, and be more efficient. This way, results obtained by others will be optimized and own results can be shared.

In order to standardize archival metadata three types of archival standards can be distinguished and used: content standards, communication standards and high-level conceptual standards. The *International Council on Archives* (ICA) has developed four international standards to describe the content of archival entities:

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

In order to communicate and to facilitate data interchange 3 communication standards are used:

- 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 Descriptive Standards). Read 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). Read 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). Read more: <https://www.ica.org/en/isdf-international-standard-describing-functions>.

⁴ ISDIAH: International Standard for Describing Institutions with Archival Holdings (CBPS - Sub-Committee on Descriptive Standards). Read more: <https://www.ica.org/en/isdiah-international-standard-describing-institutions-archival-holdings>.

⁵ Encoded Archival Description. Read more: <https://www2.archivists.org/groups/technical-subcommittee-on-encoded-archival-standards-ts-eas>.

⁶ Encoded Archival Context. Read more: <https://www2.archivists.org/groups/technical-subcommittee-on-encoded-archival-standards-ts-eas>.

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

In addition to these aforementioned standards that have been distributed in archival circles, there is another category of high-level conceptual standards that are being developed:

- the conceptual model RIC (*Records In Context*) developed by the ICA.⁸
- the *NEDA Conceptual Model* (from now on NEDA-MC) developed for the Spanish archival community. The NEDA-MC model has been developed with the intention of identifying the functional requirements that must be taken into account in the design and configuration of archival description systems. It is designed to be applicable in any environment that uses archival data, ultimately promoting the use of multiple entities and relationships with each other as a basic tool to implement current descriptive environments.⁹

NEDA-MC model and Multi – entity description

Technological changes, the development of new systems to present information and changes in the archival description method taking into account multiple agents, made it necessary to create new conceptual models for archival description, which allow a new way to understand archival science. In this respect, the aforementioned standards had to be taken into account in order to obtain a complete result. An archival conceptual model is intended to represent all the elements the archivists need to describe records, how those elements are interconnected and relate to each other, and the data we need to describe them. Both projects, *Records in Context* and NEDA-CM grew with those premises and have common elements, because the philosophy of *Records in context* is based on the Spanish conceptual model.

With the statement *we use the NEDA Conceptual model to describe our records*, is meant that we use the philosophy of the *Records in context*, or an integral philosophy to describe our records, without giving up the content and communication standards¹⁰. The use of all of the above mentioned plus new technologies and

⁸ Record In Context - Conceptual Model (Expert Group on Archival Description). Read more: <https://www.ica.org/en/egad-ric-conceptual-model>.

⁹ NEDA-CM: Modelo Conceptual de Descripción Archivística. Read more: <https://www.mecd.gob.es/cultura/areas/archivos/mc/cneda/documentacion/normas/neda-mc.html>.

¹⁰ The objective of CNEDA is to provide assessment in the standardization 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 a technical glossary (hereinafter, NEDA-Voc). Read more: https://www.mecd.gob.es/dam/jcr:c79496cf-1a09-4bb0-a977-f350b032d468/NEDA-Voc_eng.pdf.

standardization policies allowed us to achieve the principal goal explained in this article: improving access to information in an archival environment¹¹.

In order to improve descriptions in our archival systems we need to understand another concept from archival science: *multi-entity-description* and the reason why it is so important to understand this term. An entity according to the *NEDA Conceptual Model* 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 drafting an archival multi-entity description, all aspects of the document and their relationships have to be taken into account: the place of creation, the person who signed it, the people or institutions mentioned in the text, the historical event, the reason for drawing up a document, the contractor, the author, the recipient, the metadata about the origin and life of the document and the criteria of selection which were used, the reason why this record was selected for storage and which process of an institution is documented. These questions have to be answered to perfectly understand and describe the record. The context in which the document was created and used explains or completes the broader meaning of the record.

All these data can be shown in our database through entities and relationships. A document is created in a specific place and with a particular purpose, for example for a king, but surely more records were created in this place and for this king... The establishment of a relation between the description of the document in the database and the description of the archive creator as for example the king and the place where the document was drawn up (for instance the royal chancellery), makes sure users can navigate within the database and find more information related to their records. This way of searching offers the advantage to discover data that was previously unknown. This is the literal meaning of Records in Context: *The study of a document by means of the related "entities" and their representation through a computer program.*

In addition, multi-entity description offers the advantage that it is not incompatible with multilevel description; both methods are fully complementary. It remains possible to show a multilevel description as the creator of the record in the description is indicated, or as the relationship between the document and the series or the archival fonds (group of records) in which this record is contained is established. The different relationships must be recorded during the description of the records. It is for example possible to draw a relationship between the institution that created the document, the functions of that institution, other documents created by the institution, etc. In this way, multi-entity description includes multi-level description.

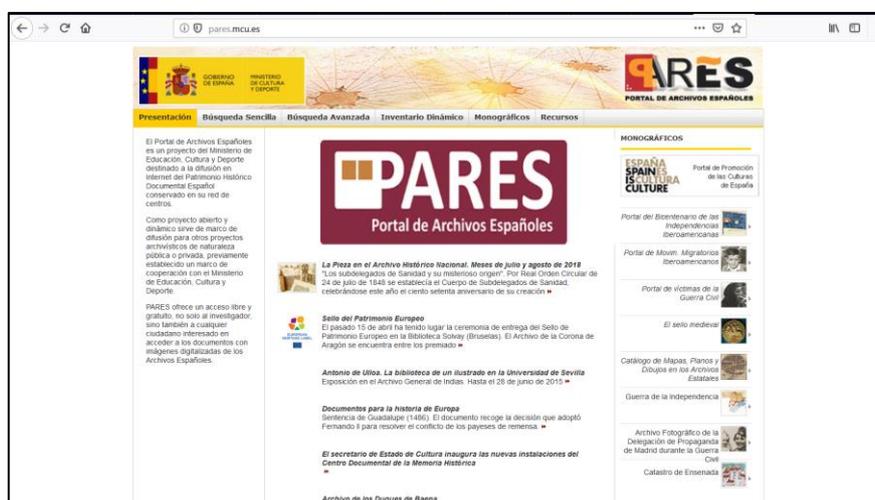
¹¹ J. Requejo Zalama, "RiC-CM y MC-NEDA:¿Nueva Descripción Archivística?" in *TRIA* 21, 2017, I.S.S.N. 1134-1602, p. 79-95.

It is also important to explain that drafting descriptions with access points is contrary to the multi-entity concept. During the description of a document, subjects or access points are in general not added. The description of a document based on a multi-entity concept goes further. It includes the analysis of the entities related to the document, the mutual relations and the visualization of it in the tool.

Describing documents while taking into account the multi-entity concept, allows us to identify, study and deal with all archival entities. Records are being kept at the center of our preoccupations. This again is the idea of *Records in Context*, and this way of understanding records allows us to grasp the full meaning and significance of our heritage, and to show it in a comprehensible way to the users.

Pares, Neda in practice

In order to carry out this task the Spanish State Archives developed the records description system PARES (*Portal de Archivos Españoles*)¹², which allows us to describe all the entities defined by NEDA CM and provides an efficient and effective access to information in a fast and intuitive way.



PARES has become the tool through which the Spanish conceptual model and its archival description philosophy was launched. The portal was conceived as a relational database, but it has been improved thanks to the application of the Spanish CM. It was created to describe records, but it has become, thanks to the efforts of many people, a tool that describes all kinds of entities that are related to each other and linked to external data. In addition, its development has taken a leap: It now captures data automatically through APIs from other cultural

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

institutions and from the different technologies developed around the semantic web.

During the description of records, archivists study them (as a primary source) while taking into account their context. All entities that are represented in the record are relevant and must be related to each other and to other records. That way, our records system grows by a considerable number of related entities and links between them. This enables us to create a real navigation effect. These interconnections will increase exponentially if we not only reuse the information but also become content providers, by converting our tool into a *Linked Open Data*-system. PARES demonstrates *Records in Context* or the multi-entity description functions in practice.

Conclusion

This article demonstrates that archival standards are not pure theory and that the way in which our colleagues have described records in previous years remains up to date and has to be taken into account. The classic archival description methodology based on catalogue cards is really useful if we take advantage of all the technical and normative tools at our disposal, thereby improving access to information.

We demonstrated that people who lack the understanding of archival vocabulary and who have no intermediary archivist at their disposal, can search information quickly, without problems, and with a tool very similar to the *Google search* function, but that makes use of the multi-entity concept and new technologies.

In Spain, we have many more issues to solve and improvements to make, but we consider to be on a good path. We must continue step by step, while always bearing in mind the users (access to information) and work in an integral way on both the record and its context. Today, PARES works fine, and citizens and researchers are satisfied.

So I would like to launch an appeal to archivists. Let us start to work hand in hand with our IT colleagues. They will understand our new archival standards very well. This way, we can make our documentary heritage known among more people, which in turn will enable us to show its value and to preserve it more easily.

TABLE OF CONTENTS

Introduction Rolande DEPOORTERE, Tom GHELDOF, Dorien STYVEN, Johan VAN DER EYCKEN.....	5
Archival Metadata Import Strategies in EHRI Francesco GELATI.....	15
Building Bridges: Preparing the Automated Transfer of Metadata for an Upcoming Data Archive in Belgium by Mapping the Metadata Standards of Archives and Social Sciences Benjamin PEUCH.....	23
Close-reading of Linked Data: A Case Study in Regards to the Quality of Online Authority Files Ettore RIZZA, Anne CHARDONNENS, Seth VAN HOOLAND.....	37
Importance of Data Standardization and Neda-Cm Implementation in an Archival Repository: Pares Ana LÓPEZ CUADRADO.....	47
The Standardization Survival Kit: for a Wider Use of Metadata Standards within Arts and Humanities Charles RIONDET, Laurent ROMARY.....	55
Behind the Scenes of Web Archiving: Metadata of Harvested Websites Emmanuel DI PRETORO, Friedel GEERAERT, Sébastien SOYEZ.....	63
Metadata and the Digital Transition: The Case of the Belgian Digital Act” Sébastien SOYEZ.....	75
Guidelines for Data Sharing and Data Citation in Social Sciences and Humanities Journals Perspectives and Insights from the Cost Action Enressh Marc VANHOLSBEECK, Tim ENGELS, Andreja ISTENIC STARCIC....	83
Data Processing and Sustainability with a Large-Scale Aggregator: the UK Archives Hub Jane STEVENSON.....	93

Digital Description and Metadata at the National Archives. Digital Strategy Jone GARMENDIA.....	103
Modern Times (Semi-)Automatic Enrichment of Metadata Eric DE RUIJTER.....	111
EHRI Vocabularies and Linked Open Data: An Enrichment? Annelies VAN NISPEN.....	117
Advanced Search Strategies in Europeana Frederik TRUYEN.....	123
Conclusion, a Vision for the Future Johan VAN DER EYCKEN, Dorien STYVEN, Tom GHELDOLF, Rolande DEPOORTERE.....	135
Personalialia.....	145
Table of Contents.....	151