Stanislava Gardasevic

Opening Archives to General Public, a data modelling approach

Supervisors: Vittore Casarosa
Carlo Meghini

Master thesis
International Master in Digital Library Learning
2011

Acknowledgements:

I would like to thank following people and institutions for their support:

My supervisor, Carlo Meghini. Thank you for putting time and efforts to tech me and guide me through this process.
My supervisor, Vittore Casarosa for supporting me throughout my work both with his valuable advices as well as with his good spirit.

The faculty of University of Parma: Ana-Maria Tammaro, Elena Corradini and Pat Dixon for their guidance at the thesis seminars.
Furthermore, I would like to thank Nils Pharo and Paul Sturges for much appreciated advices.

ISTI-CNR for providing me with the facilities for conducting the work on my thesis and especially research associate Marko Mikulicic for his great help with processing XML files.
Furthermore, I would like to thank Steffen Hennicke for advices and material from his presentation, which have made the significant contribution to the work on my research project.

My dear DILL friends who have provided me with the moral support: Danijel Cuturic, Nafiz Zaman Shuva and Nithn Lakshmana. Special thanks to Ezerea Kulisooma for help with retrieving the much needed articles for the purpose of conducting this study.

All of the professors, lecturers, and administration of the DILL master program. Thank you for giving me the chance to be a part of it. It has been a wonderful experience.

And finally European Commission that has invested in my education through the Erasmus Mundus Program. I will try my best to justify this investment.

Abstract:

By placing their descriptions on-line, Archives have gained greater public. This new public is mainly consisting of the novice users not familiar with the archival research. Archival research is conducted through the Finding Aids that serve users as a guide to the discovery of archival holdings. However those Finding Aids were originally used by the archivist for the records management and for interpreting users' requests by deriving answers from provenance and context driven descriptions. In the on-line environment, Finding Aids are usually accessible through the Encoded Archival Description (EAD) standard. The EAD was developed with the purpose of encoding and capturing many different archival descriptive practices The problem has arisen with this notion that Finding Aids in the on-line environment have the exact same form as before, just without the archivist as an mediating factor. This causes many problems to the general user public that is not familiar with the archival research process.

This thesis tends to explore one possible approach for facilitating access on behalf of the general user public to the archival holdings in on-line environment. This approach is by transforming the data encoded in EAD standard to another, more general mode. The goal model in question is the Europeana Data Model (EDM) developed for the purpose of Europeana v.1.0. project. The objective of this thesis is investigating weather EDM would bring the wanted changes to the accessibility of archival data. In order to achieve this, the general method for mapping EAD standard to EDM was developed. Furthermore the method developed was applied on the two fonds originating from the archive of Accademia Nazionale di Santa Cecilia, musical academy in Rome, for the purpose of validation of the developed method and analyzing the results of the mapping.

The results of this study have shown that transforming archival description in EDM would bring certain improvements to the non-expert users accessing on-line. The main improvements are regarding terminology, facilitated access to the different levels of the archival description, improved search functionalities and better visibility of archival holdings.

Keywords: EAD, EDM , general user, archives, mapping, access

Table of Contents:

LIST OF FIGURES:

List of Acronyms and Abbreviations:

ANSC: Accademia Nazionale di Santa Cecilia

AV fond: Audio Video Fond rom Accademia Nazionale di Santa Cecilia

AV table: Mapping table for Audio Video Fond- Appendix 5- Appendix 6

CIDOC-CRM: The CIDOC Conceptual Reference Model

EAD: Encoded Archival Description

EDM: Europeana Data Model

ESE: Europena Semantic Elements

ETN fond: Ethnomusicology Fond from Accademia Nazionale di Santa Cecilia

ETN table: Mapping table for Ethnomusicology Fond- Appendix 5

ICA: International Council of Archives

ISAD (G): General International Standard Archival Description

ISTI-CNR: Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", of the National Research Council of Italy CNR

LOD: Linked Open Data

MARC: MAchine-Readable Cataloging

OAI-ORE: Open Archives Initiative Object Reuse and Exchange

List of Tables:

CHAPTER 1: Introduction

1.1 Background

Archives have served for great number of years as a memory institutions designed to store but also to provide access to carefully selected, arranged and described documents.
For a long period, ever since the very beginning of the archival science through the great changes in the society caused by French Revolution, access to archival holdings was reserved only to privileged individuals. The situation has gradually changed and archives slowly have opened up to the public. However, with the "information explosion" caused by the Internet and the World Wide Web, archives have faced a new challenge. That challenge is to provide on-line access to archive's holdings to a new set of users, i.e. non-traditional archival users, who are not familiar with the archival research process, but used to the easy and quick access to the relevant information. Research done so far have shown that these users do not cope well with on-line finding aids, mostly because of the structure that those finding aids have inherited from the physical ones, which in turn have initially been designed to be used for the internal archival record management and to be accessed by the user with the help of a reference archivist. If these users are the target today, archives need to rethink the way of offering the information about their holdings and access to the digital surrogates of those holdings, break the wall of non-transparency and reach the new user groups.
The majority of on-line finding aids that can be accessed through the World Wide Web are encoded in a standard called Encoded Archival Description (in further text to be referred to as EAD). This standard was developed for the purpose of encoding information kept in the traditional finding aids. However, use of this standard differs in individual archival practices, which leads to a different representation of information and access possibilities.

1.2 Statement of the Problem

This thesis intends to explore whether representing archival description data in a different way could bring improvements to the accessibility of archival holdings to the general public. In particular, I intend to explore if the Europeana Data Model (in further text to be referred to as EDM), defined to describe the content of the Europeana digital library, could bring such improvements, but also if it could be a suitable representation for the description of archival holdings.

1.3 Introduction to the Archival Science and Description

Archives differ from other memory institutions in the nature of materials they have. The main difference is the uniqueness of those materials. For example, libraries collect individual published books and serials, or bounded sets of individual items. Still the books and journals held by libraries are not unique, since multiple copies exist and any given copy will generally prove as satisfactory as any other copy. On the contrary, the material in archives and manuscript libraries are the unique records of corporate bodies and the papers of individuals and families. (Pitti & Wendy M. Duff, 2001) What came to be called "archival science" emerged in the nineteenth century, and the articulation of the science's fundamental ideas, the 1898 *Manual* of the Dutch trio Muller, Feith, and Fruin, was almost entirely devoted to arrangement and description. (Wendy M. Duff & Harris, 2002)

For this thesis, it is crucial to concentrate on the description part of archival theory and practice. The definition of archival description as stated by the *Society of American Archivists* is "the process of capturing, collocating, analyzing and organizing any information that serves to indentify, manage, locate and interoperate the holdings of archival institutions and explain the contexts and record systems from which those holdings were selected". (Hensen, 2001, p. 80) Archival descriptions have to reflect the peculiarities of the archive, retain all the informative power of a record, and keep trace of the provenance and original order in which resources have been collected and filed by

archival institutions. (Gilliland-Swetland, 2001) This emphasize the central concept of archival science, which is *fond*, formalized in 1841 by the French historian/archivist Natalis de Wailly (Ribeiro, 2001). As stated in *Statement of Principles Regarding Archival Description* made by the *International Council of Archives* (ICA) the concept of *fond* implies "all of the documents regardless of form or medium, naturally generated and/or accumulated and used by a particular person, family or corporate body in the conduct of personal or corporate activity". (International Council on Archives, 1992, p.12) The most important theoretical principle based on this concept of *fonds* is "*respect des fonds*" also known as "*principle of provenance*". This principle underlines the necessity that the documents created and accumulated by a person, family or corporate body by reason of its functions or activities must not be mixed or combined with the documents of another individual or corporate body. (International Council on Archives, 1992) This fundamental archival principle is dictating that resources of different origins are to be kept separate, in order to preserve the context in which they were found and the context in which they were created. Furthermore, documents or records kept in archive are usually related to other documents, and are grouped into identifiable subgroups. This kind of record keeping and describing fosters the use of a hierarchical model. The hierarchical structure of the archive expresses the relationships and dependency links between the records of the archive. (Gilliland-Swetland, 2000; Haworth, 2001; Pearce-Moses, 2005)

Following this hierarchical structure, a fond can be organized in *sub-fonds*, which in turn can be organized in *series* and *sub-series*, formed by *archival units*, e.g. files, registers, and so on. Those units have a homogeneous nature and can in turn be divided into *sub-units* containing items such as letters, reports, contracts, testaments, photographs, drawings, etc. Following this structure of arrangement, archival description also proceeds from general to specific, as a consequence of the provenance principle, and has to show, for every unit of description its relationships and links with other units and with the general fonds. (ISAD (G), 2000) Therefore, this archival description can be presented in the form of a *tree* as it goes from the root to the leaves of the tree. This *tree* is shown in Figure. 1:

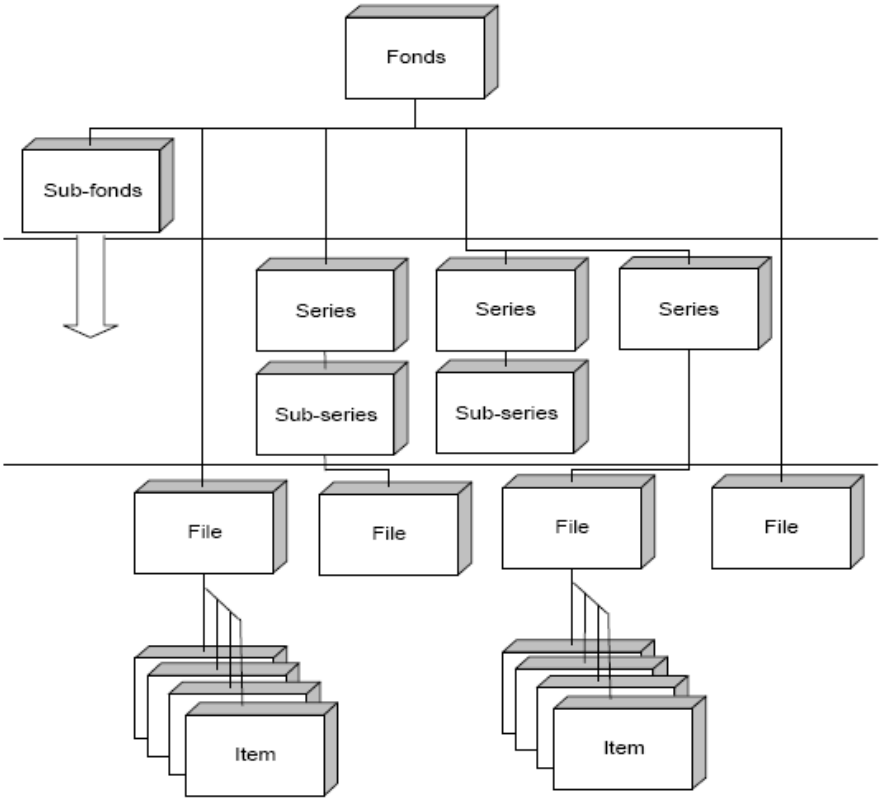

Figure 1 Hierarchical organization of the archives and of the archival description according to ISAD (G) (International Council on Archives, 2000, p.36)

1.3.1 Finding Aids

The gate to the archival holdings are *finding aids* made on archival description practice. *Finding aid* is a term ordinarily used only in archives but for present purposes it is used for all access devices that could be found in archival system, including card indexes for manuscript collections, administrative histories, inventories for archives as well as the online access versions of the same. If archival description provides accurate representation of content and context, then the user should be able to retrieve the information relevant to his or her research, and this is what constitutes the effective finding aid. (Lytle, 1980; Haworth, 2001) Finding aids are used to access archival

materials, and they contain far more information about a collection than can be found in a summary catalog record, which cannot fully capture the vast range of subjects that person is likely to find in a large collection. Finding aids are generally created in the course of processing a collection and usually reflect the hierarchical arrangement of the materials. Often, many finding aids start by describing a large group of materials, usually the entire collection or record group, and then move to the description of the series of the firs level components, followed by the description of smaller and smaller components, such as subseries, files and possibly even items. The description of lower levels inherits the description of the preceding levels.(Ruth, 2001) At the same time, finding aid acts as a collection management tool for archivist and access point for the researchers (Yakel, 2003). Yakel (2003) outlines that in the digital era, finding aids have achieved the status of having canonical form, as they are the basis for a second order representations such as MARC records, on-line HTML or SGML/XML (EAD) encoded finding aids. As this thesis deals with the EAD "canon" of finding aids, a sample of one such record can be found in Appendix 1.

1.3.2 Defining EAD Standard Through Categorization

The nature of archives, and thus of their arrangement and description, tends to reflect national, cultural, regional and organizational idiosyncrasies. On this basis ICA has developed the International Standard for Archival Description (General)-ISAD (G), and purposely left it at the general level. (International Council on Archives, 1994) However, only recently did archivists realize that the importance for sharing their descriptions lays in standardizing them, meaning crossing the boundaries of previously mentioned idiosyncrasy. Even though the constantly active discussion is ongoing between archivists when it comes to standardization of archival description, this is beyond scope of this thesis and I will not deal with this complex issue. Rather, I will try to describe the standard chosen to be the representative for archives in this thesis, the EAD standard, by comparing it to other significant ones.

Even though archival and bibliographic traditions have different approaches to identification, they still share a common suite of standards that prescribe the various components of a surrogate description and specify how that information ought to be ordered, shared and retrieved. (Fox, 2001) For this reason, Fox (2001) made a classification of these standards by separating them into four categories: **structural, content, data values** and **communication** standards and how he described them can be summarized as next:

1) Data structure standards:

these standards define the elements of information that need to be recorded about the collection, work or item. They should satisfy the answers on the questions such as: "What do we want to say about it?" or "What information is required to satisfy user needs?" (Fox, 2001, p.64) This category of standards defines not only required data elements but also the sequence of their presentation. Examples of data structure standards are EAD and ISAD(G);

2) Data content standards:

the standards that fit into this class serve to prescribe the internal form of a particular data element. For example, Rule 1.1 in Hensen's *Archives, Personal Papers, and Manuscripts* (APPM) prescribes how titles for archival collections are formed. Rule 22.5C5 in *Anglo-American Cataloging Rules* (AACR2), dictates the prescribed form of the name that will be used in a catalogue entry for the name of the married woman, whose surname will consist of her surname before marriage and her husband's surname. Another example of archival standards that fits into this category is *Describing Archives, a Content Standard* (DACS);

3) **Data value standards**

these standards support notions such as authority files and thesauri, contain list of established forms for personal, corporate, and place names, topical subject headings, and the like created on the basis of rules prescribed in content standards. In archival world

one such important standard is International Standard Archival Authority Record for Corporate Bodies, Persons and Families (ISAAR(CPF)), also developed by ICA.

4) Data interchange or communication standards

finally, this group of standards serves the purpose of establishing methods whereby descriptive data may be shared among (or within) institutions or at least among their computers. Such compatibility is obviously essential for resource sharing and user discovery. MARC format for Archival and Manuscripts Control (MARC-AMC) and EAD fall into this category. (Fox, 2001)

The previous categorization is not always applied as such in practice and it is not rare that archives use ISAD(G) as the data content standard. From the previous we can conclude that EAD standard is used for the purpose of encoding already existing archival descriptions based on data content standards and representing them online to the users, but also allowing the communication between physically dispersed archives and the union search of them. The path from the user to the archival holdings in on-line environment is demonstrated in the Figure 2:

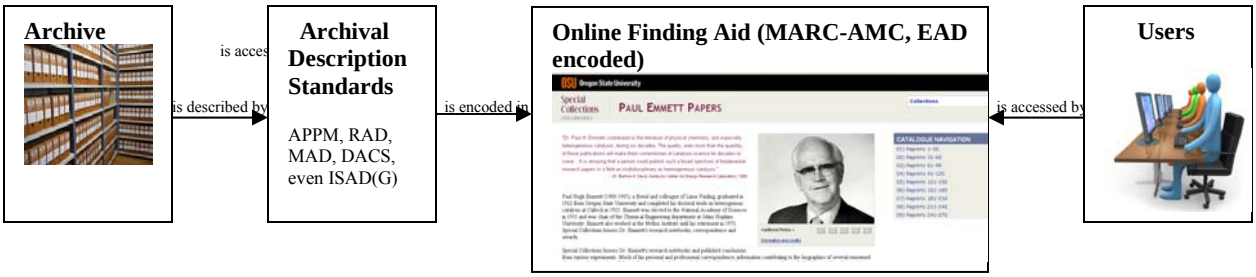

Figure 2: The path from the user to the archival holdings in on-line environment

1.3.3 Introduction to EAD Standard

The EAD standard was chosen among other archival standards because of its pervasive international use. The reason for its wide acceptance lays in the fact that it was not born

in isolation. On the one hand it fits comfortably into a generally accepted, though not fully codified, tradition of archival descriptive practice, while on the other it was compatible with new mechanism for resource discovery and delivery that were evolving on the Web. It capitalized on a suite of other standards for description that had arisen within the archival tradition but especially with ISAD(G) and MARC-AMC. (Fox, 2001) Yakel (2003) noted that the development of EAD and its relationship to finding aids can be seen as the most critical event in the evolution of finding aids to date.

Development of EAD began with a project initiated by the University of California, Berkeley, Library in 1993. This project was commenced by the need to provide networked access to the holdings of archives, but in such a way that it could include also information beyond that which could be found in the MARC encoded records that were previously used for this purpose, and found as inadequate. Daniel Pitti, the principal investigator for the *Berkeley Project* and the creator of EAD model, developed requirements for the encoding standard which included the following criteria: "1) ability to present extensive and interrelated descriptive information found in archival finding aids, 2) ability to preserve the hierarchical relationships existing between levels of description, 3) ability to represent descriptive information that is inherited by one hierarchical level from another, 4) ability to move within a hierarchical informational structure, and 5) support for element-specific indexing and retrieval."(Development of EAD DTD, para. 2). From the very beginning, the task was not to develop a data content standard , but to create instead a content designation or encoding standard able to capture all the different descriptive practices used in separate institutional practices. For this, Pitty and his team (known as the Bentley team) decided on SGML (Standard Generalized Markup Language), as the environment for their proposed encoding standard, because of its suitability to represent the hierarchical structure necessary for archival finding aids. Later on the environment became XML, as the logical successor to SGML in the Web age.

The standard (Alpha Release) was officially published in 1995 under the official maintenance of The Library of Congress Network Development and MARC Standards

Office, with the Society of American Archivists (SAA) as responsible for ongoing oversight of the standard. The EAD standard, from the very beginning of its existence has gone through several stages of thorough revisions, which included the suggestions from archival experts, implementing institutions and archival community. As the result of the revisions, several versions of the standard have been released from the version 1.0 of EAD in 1998. to the latest one published in 2002. EAD has been mapped to and from other data encoding standards such as MARC and Dublin Core. (Development of EAD DTD; Ruth, 2001; Prom & Habing, 2002).

By now EAD standard has become global. It has been implemented by a wide variety of institutions in the US, Canada, throughout Europe, Australia, New Zealand, Asia and South Africa. (Combs, Matienzo, Proffitt & Spiro, 2010) Some of the institutions using it are large scale archival gateways: For example, *A2A -Access to Archives* in United Kingdom, contains nearly 10 million records describing archives held in more than 400 repositories throughout England and Wales and dating from the eighth century to the present day. The Spanish portal *Censo Guìa* offers a structured entrance that allows navigating through the holdings of 50 000 repositories from Spain and Latin America. Other examples can be found in Italy and Germany.(Menne-Haritz, 2008) Many authors agree on the importance that this standard has to the archival community, and especially its significance for the future development in standardization.

However, this standard has been also the target of several critiques from archival theorists. Some of the critiques are concentrating on the very structure of EAD itself, while other are based on usability studies outlining the problems users have when dealing with on-line finding aids encoded in EAD. Literature review will deal with this problem in more details, together with the structural analysis of EAD and its defining schema, EAD DTD (see Appendix 2). Both of this parts are crucial for the objective of this thesis, which is to examine whether modeling archival descriptions encoded in EAD with the EDM data model can bring some improvements for accessibility to archival holdings.

1.4 Introduction to Europeana and EDM

Europeana.eu portal was launched in November 2008, as a project funded by the European Commission. The goal of this project is making Europe's cultural and scientific heritage accessible to the public. (Europeana, Background, n.d.) The first sentences of the "About Us" Web page of this portal state its main purpose: "*Europeana* enables people to explore the digital resources of Europe's museums, libraries, archives and audio-visual collections. It promotes discovery and networking opportunities in a multilingual space where users can engage, share in and be inspired by the rich diversity of Europe's cultural and scientific heritage." (Europeana: think culture, n.d. para. 1)

At this point in time, the portal is still working as the beta version, and gives access to more than 15 million objects from the various cultural heritage institutions. Those different institutions have naturally distinct ways of describing their objects, but the ones that wanted to provide their content to *Europeana*, had to conform to the interoperability solution developed for the early prototype of the portal, called Europeana Semantic Elements (ESE), which is a model based on an extended Dublin Core model. Translation to this model produced a substantial loss of data, that were recorded in legacy metadata records.(Doerr, Gradmann, Hennicke, Isaac, Meghini & van de Sompel, 2010)

Furthermore, as stated in Concordia, Gradmann and Siebinga (2009), another goal of this portal is to offer to all kinds of external communities the possibility to reuse the great amount of data *Europeana* is aggregating, by means of an application program interface (API). In addition, the technical strategy of *Europeana*, from the point of view of the functionalities it tends to offer, is to take advantage of the ever-growing Linked Open Data paradigm, by contextualizing the object representations within *Europeana* by connecting them to the already existing Web resources. Such step was, however, not possible and supported by the already existing ESE data model, as it employed a 'flat' modelling approach, which does not allow for embedding links to external resources on the Web. (Doerr et al., 2010)

In order to overcome the previously stated shortcomings of the ESE model, a new model called Europeana Data Model (EDM) was developed for the second phase of the project (Europeana v.1.0), with the launch of fully operational *Europeana* Web site with improved features and an updated interface, announced for the summer of 2011.(Europeana, v.1.0 Project, n.d.).

The work on version one of the EDM model started in May 2009, and in June 2010 version five was released. The current documentation available is regarding version 5.2. and as stated in the *Definition of Europeana Data Model Elements* it reflects the consensus reached in discussions in the Europeana v1.0 Work Package 3 meetings in 2009 and the first half of 2010 and is the result of the work of people included in *Europeana* project, but also core experts, among which distinguished ones to be mentioned are Martin Doerr from the museum sector, Michael Fingerhult from the audio-visual archives sector, Daniel Pitti from the archives sector, Emanuelle Bermes from the library sector and Herbert van de Sompel from the Open Archives Initiative. (Definition of Europeana Data Model Elements, 2011; Meghini, Isaac, Gradmann, Schreiber at al, 2010)

EDM is a rich data model developed in order to preserve original data while still allowing for interoperability. It was built not only to support the full richness of the content providers' metadata but also to enable data enrichment from a range of third party sources. "For example, a digital object from Provider A may be contextually enriched by metadata from Provider B. It may also be enriched by the addition of data from authority files held by Provider C, and a web-based thesaurus offered by Publisher D." (Europeana Data Model Primer, 2010, p.4) Even though the standard supports this richness of linkage, still it is clearly showing the provenance of all the data that links to the digital object. Also, EDM adheres to the modelling principles that underpin the approach of the Web of Data ("Semantic Web"), notion introduced by Tim Bernes Lee (2001). In this approach, there is no such thing as a fixed schema that dictates just one way to represent data. A common model that EDM brings into the picture can be seen as an anchor to which various finer-grained models can be attached, making them at least partly

interoperable at the semantic level, while retaining original expressivity and richness of original data. “EDM is not built on any particular community standard but rather adopts an open, cross-domain Semantic Web-based framework that can accommodate particular community standards such as LIDO, EAD or METS.” (Europeana Data Model Primer, 2010, p.4) In more detail, this model is discussed in Appendix 3.

1.5 Research Question

The research question this thesis tends to inform is:

- **Would transforming EAD encoded archival descriptions in EDM bring improvements to on-line access for general archival public?**

1.6 Justification of the Study

This study was designed with the ***aim*** of facilitating access to the archival records in on-line environment to wide general public. In her essay *Access – the reformulation of an archival paradigm,* author Angelika Menne Hariz (2001) has stressed that the focus of archives is shifting from storage to access. Among others, she raises the issue of the necessity for designing a full range of new instruments and concepts, that provide orientation and help to find the way to the material that can deliver the information needed. As Albert Einstein stated: “We can not solve problems using the same way of thinking that created them”, this research is not following the direction of investigating possible improvement of access to the general user by concentrating on the EAD model itself or the interface design issues that should be developed upon this model. The path chosen for this, in essence prescriptive research, is by **modelling the data** encoded in EAD through a new approach that EDM brings towards structuring and representing this data. This model was chosen because it was developed on the new technologies that allow accessibility improvements, but also because it was built upon established standards, through consensus process with the direct contribution of cultural heritage experts, including archival experts. Therefore, the ***objective*** of this research is examining

weather EDM would bring the wanted changes to the accessibility of archival data. Furthermore, this study would help to identify whether this model is supporting adequately this archival legacy schema and if the benefits that it may bring to the discovery of archival data are bigger than the possible data loss that the conversion process might carry with it. In order to examine the possibilities that this standard may bring to accessing archival data, a method needs to be developed for translating the data from the source model (EAD) to the goal model (EDM) and further applied on the real data in order to examine the changes.

1.7 Research Design

As previously stated, the main problem this thesis is addressing is examining possible improvements that the EDM model could bring to representing archival information encoded in EAD to the general public accessing archives on-line. The problem is going to be addressed in several steps.

-what do the members of general user population find problematic with current on-line finding aids, and what are their preferences for access;

-how to develop general mapping method from EAD to EDM;

-validating this method with a real life case (data from archive of *Accademia di Santa Cecilia* in Rome);

-analyzing the results of the validation to see to what extent the newly created EDM data can answer the issues identified in the first step.

Therefore, to provide the basis for this study and try to find an answer to the first point, I will review the literature dealing with users and usability studies. This literature mainly deals with on-line archival finding aids EAD encoded, but where results may be relevant for the purpose of this issue, results of user studies on traditional finding aids will also be included. Further literature regarding the structural issues of EAD, that may have influence on usability will be included as well as relevant content analysis studies. Out of this body of literature and recommendations from the researches, the Theoretical

Framework is to be developed that should present a set of issues found crucial for the non experienced archival user accessing archives on-line.

Guided with this theoretical framework, a method is to be developed that would provide the mapping solution from the source (EAD) to the goal (EDM) schema. It is necessary to stress that EAD schema is consisting of 146 elements and numerous attributes that specify the metadata that those elements may hold. Furthermore the use of the elements is idiosyncratic in different institutional practices, therefore for the purpose of this research, which is at master level and time limited, I will not be dealing with the whole schema, but only with a subset of it.

The next step would be to validate the method previously developed by applying it on a real case, which is the EAD data provided by the archive of *Accademia di Santa Cecilia* in Rome. This archive is holding mainly multimedia materials, and the data I will be applying the method to will be the EAD encoded description of two fonds: *Audiovideo* and *Ethnomusicology* .

Based on this conversion, analysis of the original data and the newly modeled data will be done having in mind the questions and problem areas raised in the theoretical framework, coming from the literature review, in order to answer to the research question.

1.7.1 Limitations

Inherent in the design and implementation of any study are certain limitations. The methodology chosen for the purpose of this research project allows indentifying some of the improvements that transforming EAD to EDM will bring to the general user population, however these conclusions are made on theoretical basis. In order to confirm these findings some additional research should be conducted. However due to the time limitation of this thesis and lack of interface solution for EDM remodeled archival data,

this was not possible. Suggestions for such further research can be found in section 5.3 of this thesis.

Furthermore, the validation of the general method for mapping EAD to EDM developed in this thesis was conducted on two fonds from archive of *Accademia Nazionale di Santa Cecilia*. However, these fonds may not be the best representatives of EAD encoded archival description, considering the lack of some data elements that are usually used in archival practice. Still, they have served the purpose of applying the method and analyzing results sufficiently in order to answer the research question. The justification for choosing these specific fonds and some further limitations concerning the Validation work are stated in Chapter 4.

1.8 Outline of the Thesis

Chapter 1: Introduction provides the background of the research topic and statement of the problem. This is followed by a short introduction into archival descriptive practice, the EAD model and the EDM model. It includes the research design chosen for this study, research question, limitations, and outline of the thesis.

Chapter 2: Literature Review is dedicated to exploring access to the on-line archival aids and the problems non traditional users may face with it. Also it tends to conclude what are the users' preferences. It is reviewing EAD inherited problems, user and usability studies done so far but also it looks at the problems indentified within the structure of EAD. It is built in a form of a Theoretical Framework that should help conducting this study. The important parts of this literature review are also structural overviews of the two models in question, EAD and EDM, and can be found in Appendices 2 and 3, respectively. Those two chapters are considered as crucial for understanding the work presented in Chapters 3 and 4.

Chapter 3: Research Design describes the methodology chosen to implement in this study and provides a justification for that choice. In this chapter a general Method for mapping EAD to EDM is developed

Chapter 4: Validation and Discussion will apply the method developed in Chapter 3 to the two fonds from Accademia di Santa Cecilia. Further it will discuss the results of the mapping by comparing the original data and the newly modeled data in the perspective of the general user access problems and preferences indentified in the literature review, in order to investigate the possible improvements of this translation.

Chapter 5: Conclusions offers conclusions to the research question presented at the beginning of the thesis. It provides with the further remarks that EDM brings as a possibility to archival data discovery. Finally, some suggestion are made for the directions that might be taken by future researchers focusing on this topic.

1.9 Chapter Summary

This chapter has provided with the foundation of the thesis. The problem the thesis is dealing with was introduced and necessary introduction to the archival description practice, EAD standard and EDM model was provided. Further, research question was introduced, followed by the justification of the study, its research design and limitations. Finally, the outline of the thesis was provided. The next chapter, Literature Review, is going to build up on the issues introduced in this.

CHAPTER 2: Literature Review

This chapter will review the literature that helps to identify key issues for general users when accessing archival finding aids on-line. These key issues will form a theoretical framework that should serve as a guide when developing a method for mapping EAD and EDM schemas. This is crucial for the concluding part of the thesis as it will provide criteria through which I will discuss the possible improvement that EDM-remodeled EAD data could bring to the general user population.

The literature reviewed for the purpose of developing this theoretical framework is mainly about EAD standard and user studies on on-line finding aids encoded using it. Because of the lack of user studies conducted specifically on EAD encoded finding aids, other user studies dealing with the archival research process were included. The other literature found to be relevant, and therefore included in this literature review, deals with the EAD model itself, its structural analysis and content analysis of EAD encoded finding aids.

In order to consult the best sources for the search of relevant literature for this study, I used the results of an evaluation done by Australian Research Council Research Excellence that examined peer-reviewed journals from the field of Archive and Record Management. (The complete list of journals in all disciplines is available for download as an Excel spreadsheet at URL: http://www.arc.gov.au/era/era_journal_list.htm). Based on results of this evaluation I have searched for the literature within the top ranked journals, such as: *Archival science*, *Archivaria* and *American Archivist*. Furthermore, with the literature retrieved I have applied the information search strategy of *footnote-chasing* (Bates, 1989). Additionally, searches of the literature were conducted in the JStore data base and World Wide Web, especially using Google Scholar.

Terms I have used for conducting the searches include various combination of the following: “EAD” “user study”, “usability”, “user”, “finding aid”, “archive”, “metadata”, “standards” “internet” “on-line”.

2.1 Introduction to the Theoretical Framework

In *Encoded Archival Description on the Internet* which is dedicated solely to this standard, (Pitty and Duff, 2001), only one article deals with delivery of EAD encoded archival description to end users. The author of this article, Gilliland-Swetland (2001), questions the functionality of finding aid as an information discovery and retrieval tool, by stating that archivists have historically been materials-centric rather then user-centric in their descriptive practices, which causes problems for the users. The author has outlined several key issues that have served as a model and inspiration to the theoretical framework I have further developed in the literature review. The problematic areas found in this paper I have summarized as following:

- lack of alternative access points for users, because of the arrangement of materials according to provenance or original order of records. The author emphasizes the importance of subject access. Furthermore, the technical language of creators of materials and archivists causes further problems for discovery. The author argues that archivist needs to map technical terminology used as *subject access points* and for *labeling data elements* to a less technical vocabulary in order to facilitate resource discovery by non-expert users; (Gilliland-Swetland, 1998)
- finding aids consist of extensive contextual description of the circumstances surrounding the creation of its materials, when not all users or uses require or desire contextualization. However, the traditional finding-aid makes direct, de-contextualized access to archival materials close to impossible. Length of the files and navigational complexity makes the process of discovery even harder. Also, users find the administrative information that is woven throughout finding aids confusing;

- lack of item-level description, problems caused to users by describing material collectively and hierarchically since they want item-level access as well as locating a known item quickly. The author states that some users may want to invert the hierarchical method of information discovery, and start from an individual item in order to investigate its context and search for other related ones.
- the traditional finding-aid is designed to be used in an environment where archivist acts as a mediator between user and the finding aid. On the internet there is usually no such possibility. (Gilliland-Swetland, 2001)

These issues are still current, but since 2001 there have been more studies regarding users and the usability problems of on-line finding aids. Thus, for the purpose of my work, issues outlined by Gilliland-Swetland will be further developed in a theoretical framework that should identify key problems the general user faces during on-line archival research, but also deals with suggestions that authors have made in order to overcome them. This framework is mainly made up of the studies found relevant to this question, published in the last 10 years. It is important to note that the user studies conducted in controlled settings on EAD encoded finding-aids I have surveyed come predominantly from the North American continent, while those originating in Europe (UK to be specific) deal with log analysis of institutions giving on-line access to their holdings descriptions.

2.1.1 Usage and Users of Archives

Usage of archives has drastically changed with their presence on-line. Archival research is a very precise process, and because of its complexity, archival science theorists, such as Richard Cox, questioned whether online finding aids would even find an audience apart from other archivists. (Cox, 1998). However, recent researches have shown differently. Online users have become much more numerous than those who approach archives physically.

The State of State Records, a status report on state archives and records management programs in the United States (2007) has shown that the number of users making direct, person-to-person contacts with staff, has grown slightly over the last decade, however their methods of contact have changed dramatically. While reference request received by email have been increasing, other forms of person-to-person contact have declined since 1994, e.g. surface mail (down 24 %) in person visits (down 17 %), and telephone calls (down 13 %). On the other side, number of users of web sites has grown immensely. While the exact number is not known, authors of the report have made a safe guess from the available statistics that there could easily be 100 Web visitors for every user who requests assistance or information through a person-to-person contact. This number is ever growing, as comparable data on unique visitors from years 2004 to 2006 have shown growth of 109 %.

Next question is: who are those ever growing users of archives?

Conway's research at the U.S. National Archives, identified four major categories of archival users: *academic* (e.g. academic historians, humanity scholars), *occupational* (e.g. institutional administrators and K-12 teachers), *avocational* and *personal*. In this report, genealogists were placed in the group of avocational users, but in reality, depending on the level of the genealogical research they are conducting, their work can be seen also as professional or personal.

The research previously mentioned in USA State Archives also examines who are the most prominent users. The conclusions drawn are that group of genealogists are the single largest constituency of users for state archives, making more than 50% of total users The second largest group of users access the state archive for the purpose of administrative usage, for state or local government, while the third place is taken by users who access state archive for the purpose of property or legal research. The next groups are scholarly and researchers of local history and the smallest percentage of usage was recorded by the groups classified as "other", undergraduate students and people who used

archives for purpose of "kindergarten to twelve grade" (K-12) projects. (The State of State Records, 2007)

The results found in USA correspond to the ones found in Europe, UK in particular. As part of an investigation of the requirements for online searching software, under the project called *LEADERS,* research was conducted into the searching facilities required by different types of users. Six archive repositories took part in this survey of users: *The National Archives*, the *Wellcome Institute*, *Dorset Record Office*, *Birmingham City Archives*, *Glasgow University Archive Services* and *University College London's Special Collections*. This research found that the majority of users (60 %) fit into the category of personal leisure use. The next largest grouping (at 22 %) was made up of people using archives as part of their job, which includes academics and professional researchers. Furthermore, this project went on to analyze these users' subjects of research: 64 % were interested in looking for information about families, individuals or organizations, while 23 % were looking for a particular topic. Also, the correlation between these two sets of groupings was high, as 84 % of personal leisure users were looking for families, individuals or organizations and 85 % of professional and educational users looking for topics.

2.1.2 Who is the "General" Archival User?

It can be clearly concluded that archives have gained a new public that is made of users arriving from descriptions obtained on-line. In the essay *Online Finding Aids: Are They Practical?* Hostetter (2004) states that by putting finding aids online, repositories invite a wider audience to access their collections and expand their clientele beyond 'scholarly researchers' to include members of the general public. This is exactly the problematic area mentioned by Cox (1998) at the beginning of previous chapter. By questioning the very usability of on-line finding aids, he has expressed the concern that the average researcher, member of that general public, would be unable or unwilling to browse a finding aid without the assistance of an archivist to explain archival concepts or to guide the researcher through the occasional complications of these finding aids.

For the purpose of this research the users that will be addressed as **general**, **non-traditional** or **novice** or **non-expert users** are those that are accessing archive on-line, but not possessing so called "a*rchival intelligence*". Yakel and Torres defined *archival intelligence* as:

> *"a researcher's knowledge of archival principles, practices, and institutions, such as the reasons underlying archival rules and procedures, how to develop search strategies to explore research questions, and an understanding of the relationship between primary sources and their surrogates."*(Yakel & Torres, 2003)

Prom (2004) adds that archival expertise or intelligence is gained by using archives and conducting historical research, either as an archivist or as an experienced user. Studies have shown that the archival researchers with higher levels of experience in using libraries, online library catalogs, and archival finding aids have greater search success than the novice ones. (Daniels &Yakel, 2010) Authors agree on the notion that offering online access to archival holdings very much changes the process of research in archives. It requires revising the archival practices for offering access and meeting the challenges so the user needs can be addressed more fully. (Yakel, 2002; Prom, 2004; Giliand-Sweetland, 2001)

The fact is that online users are numerous and that their number is actually much bigger than the number of direct users of physical archives. The few studies conducted so far on these new users have shown that they indeed encounter problems when accessing archives on-line. Before concentrating on the problems users encounter and recommendations for solving them, the next chapter will introduce the theorists' view on the EAD model itself and what may cause these problems.

2.1.3 Problems with EAD Encoded Archival Description

This section deals with the problems that are intrinsic to the EAD model itself, that may be affecting the on-line users who are accessing archival holdings through it, as it was shown on the Fig. 1. EAD standard is a product of the archival profession, and in order to understand the problems of this particular standard, it was considered necessary to explain the practice and philosophy of archival description behind it, in the Introduction chapter.

Eidson (2002) in his critical essay with an interesting title *Describing Anything That Walks: The Problem Behind the Problem of EAD* brought to surface the problems that are in his words "built into very fabric of EAD" (Edison, 2002, p.7). This author emphasizes that the main problem is the fact that EAD is used to encode exactly the same information found in traditional finding aids. The assumption was that traditional finding aids were adequate resources to help researchers find what they need. Authors have been suggesting that archivists assume too much about their users, when it is often the case that those assumptions are wrong. (Conway, 1994; Dearstyne, 1987).

Furthermore, the finding aids did not only serve for the discovery purposes of end users. They have served archivists over the years as effective tools for the management of their collections. (Landis, 2002) Proof of this practice is apparent in common processing terminology, such as "unit title". Consequently, as the on-line versions of finding aids copy the physical one, they copy also the practice applied to the physical one, which is in turn confusing for the non-experts.

Furthermore, the traditional finding aids were designed to be accessed with having the reference archivist as a mediating factor. As many as 85% of users rely on the archivist to do the research for them, especially when interpreting subject or name-oriented requests and deriving answers from provenance and context driven descriptions. (Pugh, 1982) Still, EAD finding aids are organized on the same archival principles and contain the

same content as traditional finding aids, but in the on-line environment and without the help of a reference archivist, the discovery of primary sources can be hampered.

Finally Eidson (2002) points out that EAD wasn't built with the user on the mind, but was built by archivists for archivists. During the first few years of implementation of EAD authors have shown concern that ignoring users would cause problems with EAD development and its promotion and acceptance by archivists world-wide. Furthermore, it was outlined that none of the EAD primary resource material discusses the effectiveness and usefulness of EAD deployment methods from the point of view of the end-user. (Edison, 2002). On the other hand, Coats argues that this lack of user studies in EAD has been a possible determinant of the widespread acceptance of this standard among archivists. (Coats, 2004)

Consequently, all of the previous points have influenced the general user's experience when accessing on line finding aids. The framework developed in what follows deals with these issues in more detail.

2.2 Theoretical Framework

As already stated, the theoretical framework developed for the purpose of this study examines the problems of general users accessing on-line finding aids and outlines suggestions from the literature for overcoming these problems and facilitating on-line access.

As opposed to libraries and their digital counterparts, archives don't have a long tradition of conducting user studies in order to examine how their public actually behaves when searching for information. The cause of this may be that many archivists believed they had an "instinctive sense" of what researchers needed. However, it is questionable whether or not archivists truly know their users so well. (Maher, 1986; Landis, 1995). This point is emphasized by the fact that development of web sites makes information about archival repositories and items from collections available to a dramatically larger,

virtually unknown audience, creating the need to become familiar with these "unknown" archival users. (Coats, 2004)

Still, while searching for relevant literature for this chapter I have noticed that the number of published user studies on-line finding aids is still very scarce, at least in the English language in which the search was conducted. Most of the studies will be reviewed further in the following sections.

2.2.1 Issues Regarding Archival Terminology

The most prevalent problem in the literature reviewed is regarding to the terminology used in on-line archival finding aids. A common complaint from the users is about misunderstanding or misinterpreting terms that are ambiguous and have their roots in archival parlance. (Duff & Stoyanova, 1998; Yakel, 2004; Scheir, 2006; Johnston, 2008; Daniels & E. Yakel,,2010) Both experienced and inexperienced archival users face difficulty with this archival terminology. (Prom, 2004).

In a pilot survey conveyed at the University of Pittsburgh, Yakel (2004) brought together a representative focus group of potential users, of archives to determine the practicality of online versions of finding aids. The users involved were mainly graduate students with similar backgrounds, who had little previous experience with using archives. One of the findings of this study was that the archival jargon used throughout the finding aids made difficulties for the users. The terms that caused most ambivalence such as **abstract, scope, content notes, historical sketch,** and also **abbreviations** (e.g. TLS-typewritten letter signed) all originate in archival terminology, which is unfamiliar to the common user.

Similarly, another study conducted on novice archival users found that the specialized archival terms or their variations that were most confusing to the users in the sites accessed for this study were: **Finding aid**, **Creator, Extent, Repository, and Scope and Content**. (Scheir, 2006) Another term also found to be confusing was **Index**. (Johnston, 2008) In

order to understand the terminology used, one would have first to understand the underlying concepts and meanings, which the novice user does not (Yakel &Tores 2003). Schier's (2006) results have suggested something different, but only in cases where other aspects of the site were designed to facilitate user searches. In these cases full understanding of the archival terms was not necessary in order to move through a finding aid in a meaningful manner.

On top of the confusion that using archival terminology creates, what only intensifies this feeling, is the inconsistent use of this terminology between different institutional on-line finding aids. The terminology differs in terms of both their meaning and wording. For example, encoders labeled the *physical description* data element (<physdesc>) using three different terms: *size, quantity and extent*. Even though the wording of these terms was different, their meanings were almost the same. On the other hand, for encoded *acquisition information* data element (<acquinfo>), several different labels were used, including: *acquisition*, *date of acquisition*, *immediate source of acquisition*, *acquisition information*, *provenance, donor, origination, and acknowledgement,* among which "provenance," "origination," and "acknowledgement" were even not appropriate terms for the presentation of acquisition information. If the meaning of labeling terminology is not consistent with the content, users will get even more confused. (Kim, 2004)

2.2.1.1 Recommendations Regarding Terminology

Findings suggest that defining, clarifying or eliminating certain archival terms, while still retaining their categorical integrity would result in their better understanding by novice users. For the display of the finding aid in an online environment, confusing terms and phrases might easily be translated into terms and phrases that make more intuitive sense to general users. Shier (2004) gives example of *Scope and Content*, which is an important section in many finding aids describing the bulk, content, extent and limitations of a collection, which is however a specialized, unnecessarily opaque descriptor. The author suggests that it might easily be converted into a phrase as straightforward as, "What is in this Collection?" Another such term is *Arrangement*, about which one of the participants in the study thought to be "an unclear way to indicate the organization of the information." (Shier, 2004, p.63) The author

suggests that the reduction of coded terminology enhanced users experience with the site, as it was the case for the one that used Arrangement to organize its material, but the exact term never appeared. Prom (2004) agrees with this by explicitly stating that archivists should just try avoid the use of archival terminology.

To prevent language-based confusion, except for using clear and precise language, Johnston (2008) suggests that those terms should be standardized across finding aids to make them more accessible to repeat users across repositories. Kim (2004), adds to this point by stating that since EAD is a data structure standard, it needs a corresponding data content standard in order to present information more consistently.

2.2.2 Issues Regarding Structural Representation of Finding Aids

In order to reach the desired material or information, archival web sites usually offer the possibility to navigate through finding aids, following the hierarchical structure in which these materials are placed. More about the browsing function per se, will be discussed in the next section. This section deals with users' behavior when encountering this top down path to access document descriptions.

An important issue that causes problems to general users, is the mode of organization, description and consequently the presentation of finding aids. Finding aids are generally created in the course of processing a collection and usually reflect the hierarchical arrangement of the materials (Ruth, 2001). This term "hierarchy" denotes the structure of a finding aid whereby multiple levels describe a collection's arrangement. In many finding aids this structure is represented by using an outline form showing a progression from a general description of how the collection is organized (*Arrangement*), to descriptions of each record group or series (*Scope and Content*), to the Container List of folders within each series (and sometimes subseries).(Scheir, 2006)

As previously mentioned such types of organization and description of documents are considered to be crucial for archival science as they allow sufficient contextual

background during the discovery of relevant material. Menne-Haritz (2001) states that archives provide information potentials, not the information itself, and that they cannot be read, but understood. They enable an investigation, which is possible only with the rich contextualization. However, studies have shown that while this type of description and access to documents is preferred by more proficient users, it causes problems to the novice ones. Novice users have little understanding of provenance or the context in which collections might be assembled and in some cases are not interested in having this rich contextualization (Daniels & Yakel, 2010, Gilliland-Swetland, 2001)

These results were apparent in a study conducted by Prom (2004), that involved 89 individuals in interaction with University of Illinois archives and its online access. The individuals were further divided into 3 groups consisting of self-proclaimed proficient archival users, proficient computer users and novice users. The study has shown the gap in search success between the proficient (the results have shown that archival and computer experts had similar results) and the novice users. While the experienced users prefer completeness of the archival description and browsing through the fullest finding aids in order to find relevant material with its full length and complexity, novice users expressed the feeling of being lost and "did not know where to begin searching". (Prom, 2004. p. 25.)

This result was repeated also in Yakel's study (2004) where subjects also claimed they "got lost" in the hierarchy of the finding aid. This feeling was especially present within the full text view of the finding aid. Big chunks of text that may appear in finding aids made them impatient because the need for extensive scrolling. Similarly, the study by Daniels and Yakel found that characteristics such as large blocks of text and hierarchical presentation pose special problems for searchers. (Daniels & Yakel, 2010)

The previously mentioned study by Scheir (2006) examined user access via six online finding aids chosen to represent a range of characteristics common to many online finding aids. This study involved novice users (non academic and non historian), with the addition of one archivist who was invited to join the study in order to compare results of the

search with the novice users. The author states that the structure of many online finding aids is built on the assumption that the multilevel, contextual environment of a finding aid is self-evident to its users, but the findings have proven different. The subjects have shown inability to make conceptual connections in order to navigate through them (e.g. lack of understanding of the relationship between series and boxes). This issue brings questions of terminology and archival structure together. Those not familiar with the meaning of *Arrangement* in the archival context or with the hierarchical relationship between *Series* and *Container List*, often could not make sense of the site structure. This is clear especially when compared with the performance of the one Archivist involved in the study, who had prerequisite knowledge of the arrangement of material, and was moving through the finding aids with ease, able to find her way to the answer much more efficiently than the novice users.

However, it has also been noted that participants experienced a learning curve during the experiments. They reported greater confidence and more ease in using the system as the studies progressed. (Sheer, 2006; Johnston, 2008). Still, these studies were conducted in controlled settings. When it comes to the real life situations, one can not help but wonder: 'How much patience or willingness to learn can archivist expect from the novice user who access archive on-line?' In online finding aids it is often presupposed that users either have a sound knowledge of the series system and the terminology used or are prepared to acquire it, but such presuppositions potentially hinder new users from fully understanding, enjoying, and exploring archival collections through their online descriptions. (Rosenbusch, 2001)

2.2.2.1 Recommendations

Prom states that the less hierarchy online finding aids have the better (as stated in Scheir, 2006). However, many archivists take a traditional finding aid structure to be an indispensable descriptive analog to the collection it represents, and thus consider such structure critical for maintaining intellectual control over archival collections. Further, archivists contend that multilevel displays guide researchers to understand a collection's

provenance, a way to impel an understanding of informational content within the context in which it was created. For this reason, many believe that online finding aids must be structurally identical to their print versions, and therefore are likely to resist Prom's recommendation to reduce hierarchy. (Scheir, 2006)

Kim (2004) like Prom, observes that hierarchy in online finding aids causes confusion for some users. But where Prom suggests that multiple levels might best be minimized in online finding aids, Kim points out to a study by Altman and Nemmers that found that "users should know where they are in the collection at all times."(Altman & Nemmers, 2001 p.126.) concluding that online finding aids should retain their multiple levels, but must also include design elements that heighten user awareness and understanding of the significance of such structure.

2.2.3 Issues Regarding Accessing Possibilities

Researches have shown that many archival web sites do not even allow a search function on their EAD encoded finding aids, even though this would undoubtedly be an effective way for users to access finding aids. Nevertheless, in such systems access is possible only through a browse function or limited search function that allows users only to perform full-text searching within the structured finding aids.

The dominant search strategy when faced with on-line finding aids is either: scrolling through the finding aid or using the browser's find-in-page (CTRL-F) function. (Prom, 2004)

Electronic finding aids tend to consist of a single or only a few web pages that are very long and filled with text. In some cases web sites provide "Search in the text" functionalities, while in other Ctrl+F browser's find-in page function is extensively used by users when they are faced with large blocks of text. (Bantin, 2001; Prom, 2004). Users confronted with large chunks of text and no pictures get intimidated and frustrated by the necessity for extended scrolling. (Johnston, 2008)

The first impulse of the users is to get access to material through a search. This was recorded by Prom (2004), whose study shown interesting and contradictory results. When asked, the biggest percentage of the participants in the questionnaire expressed a preference for accessing the descriptions of materials by scrolling through the website and clicking the links, and a much smaller number claimed to prefer searching by keywords. However, contrary results were obtained when participant behavior was actually observed and coded. Most users used a search box when one was available, and only six of the 26 who claimed to prefer browsing or clicking links actually used that as their primary technique. If users do not exactly know the title or the creator of the collection that they want to find, browsing can be time-consuming. Searching should be the more effective way for users to access, especially as the number of encoded finding aids on a site increases. (Kim, 2004).

Shier notices that her subjects have expressed "desire to obtain immediate answer, with little passion for following steps down a hierarchical path intended by the representation to put the information in the context". (Schier, 2004, p.60)

The web sites providing search functions usually offer only simple search, which can, in some cases, be delimited by different encoded elements. Fewer offer advanced search options. In a study conducted by Zhou (2007), among fifty-eight web sites examined, forty-five employed a search system, fifteen of which had both simple and Boolean search options and the rest employed only a simple search.

2.2.3.1 Search Parameters

For facilitating access to archival records it is of crucial importance to know how users want to access information. This information should be used both by those who make the description and encode this description as well as by the implementers of search engines and designers of interfaces. For the purposes of this study this is also important in order to utilize the EDM in the best possible, so that it can serve as, in a way, a query language for accessing archival holdings.

The current situation can be observed through two content analysis studies (Kim, 2004 and Zhou, 2007) that pointed out some of the parameters currently used by archival web sites. Kim (2004) notes that about 30% of the web sites examined provide search parameters, such as: repository, names (personal/geographic/cooperate/family), places, subjects, call number, collection title, scope and content notes only, library, catalog headings, front matter, container list, and full text.

In the other study done by Zhou (2007), out of total number of forty-five archival web sites that provided search options, an occurrence frequency of search parameters is shown in the Table 1 below:

Ranking	Options for Searching	Frequency
1	Full-Text/Keywords	38
2	Collection Title (or Title)	21
3	Subject	15
4	Inventory/Repository/Library	7
5	Scope and Content/ Biography/History	7
6	Author (Family, Full Name)	7
7	Names (Family Names, Full Name)	6
8	Geography Name/Places	6
9	Call Number	6
10	Descriptions	5
11	Dates	4
12	Creator	3
	Abstract/Summary	3
	Container List	3
	Others	≤ 2

Table 1: Frequency and Ranking of Options for Searching (Zhou, 2007, p.108)

The statistics of the current users and usage of archives show that the most important access points are *subject* and *name* parameters. (Hill, 2004) This means that current state of on-line archives providing search possibilities is not at the appropriate level, and as we can see in the Table 1 above, only 6 out of 54 institutions that have EAD encoded finding aids allow their users a *name* search, while a *subject* search is better supported.

For facilitating access to on-line discovery of archival holdings, the crucial thing is to know what the users want to have access through, and what the information is that they want to pursue, except for the already mentioned *subject* and *name*. Most of the studies in

this literature review evaluated so far, deal with user behavior with on-line finding aids. However, these researches are mainly done in a laboratory setting with prepared and predefined search terms and tasks.

The results of the LEADS study, previously mentioned, which examined the actual usage of on-line archival systems in UK emphasized the importance of providing both detailed finding aids (which record the names of individuals and locations) and guidance to the subject strengths of collections (Hill, 2004). For the purpose of further investigating the most important parameters for searching, the most relevant aspect is user studies done on the representatives of groups of archival users, in order to track their real search behavior and note their access needs.

2.2.3.1.1 Genealogist

As already noted, the most numerous users of archives are researchers interested in exploring family history, i.e. genealogists. Genealogists are predominantly seeking and finding facts about people. They do so by searching in the first place for the *names*. *Names* that are of interest for them are mainly *personal names*, but searching by personal names provides a number of challenges, including the need to differentiate between people with the same name or to retrieve names with different spellings. The other important access point is *place names*. The need to search by *place* to locate information about people was emphasized by almost all the participants in study conducted by Duff and Johnson on the user behavior of genealogists in physical archives. Another point emphasized was that the *names* and boundaries of *places* often change over time and archives usually organize and index records by the name a locality had when the records were created. Therefore, for genealogists it is crucial to be able to link current place names with former place names, and vice versa. The next highly important parameters for this group are possibilities to access through delimiters of the specific *genre* of the document and *event* to which the document is related, and to limit the search by the *date*. (Duff & Johnson, 2003)

The results of the study by Duff and Johnston (2003) have also shown that genealogists have learned to work around the archival systems because the systems do not meet their needs. Not surprisingly, researchers came to the conclusion that novice genealogists find provenance-based finding aids confusing and frustrating to use.

2.2.3.1.2 Occupational Users

The second most numerous group of users of archives are the ones using them for occupational reasons. Gilliland-Swetland (2001) has separated two groups that belong to this sort of users. First are *institutional administrators*, who work for the institution that created the archival material. For the purpose of successful completion of their work, the author has suggested several access points that would be of use to them, such as: key *events*, that could be chronologically presented, the possibility to search by *genre*, *date* and *table of contents*, or by *format* (e.g. galleries of institutional images, searchable by *subject*).

Secondly, a small but identifiable group of users of archives use them for the purposes of primary and secondary education projects (K-12) (The State of State Records, 2007). The study conducted on the K-12 teachers, found that this group's interest is seldom context, but they wish to locate and contrast specific items from several collections, exhibiting characteristics that are representative of a ***genre, format, period*** or ***event***. (Gilliland-Swetland, 1998)

2.2.3.1.3 Academic users

The conclusions elicited from the researches done by Tibbo (2002) with academic historians and Bates, Wild and Siegfried (1993) with humanities scholars demonstrate that while for this particular group of users provenance and wider context of the materials is crucial for their research, access point such as *dates*, in chronological order, *geographical locations*, *names* (individual and group), *discipline terms* and *topical subjects* are still very useful. (Gilliland-Swetland, 2001).

2.2.3.2 Recommendations Regarding Access Possibilities

The following recommendations should be taken in consideration both by archivists doing the actual descriptions of archival holdings and by implementers of on-line discovery systems. What can be summarized from previous chapters is that it is crucial to allow the users not only to browse, but also to search in finding aids. In order to allow good search functionality it is important to have quality metadata assigned to the records. Archival description, however rarely goes to the lower levels of material, even though authors agree that for some users it would be crucial to have access at the item level , and all the users would benefit from this. (Gilliland-Swetland, 2001; Hill, 2004; Kim, 2004)

The parameters for searches that would be most useful are in the first place, *subject* and *personal name*, followed by *geographical location***,** *genre* and *event*. Proof of the utility of allowing discovery through *name* can be found in statistics showing that the most frequently accessed catalogues in *Access to Archives* portal of UK national archives (A2A) are those for "Quarter Sessions", which contain large numbers of names of individuals. (Hill, 2004)

Also, since users tend to have difficulties in selecting search terms it is highly recommended to provide controlled subject indexing and access. Studies have shown that a difference exists between recall and known-item searches. Recall searching often benefits from synonym generation, identification of controlled access terms, and knowledge of the topic which users may not possess. To assist with recall searches, online finding aid systems might be modified to suggest related terms to searchers, or to provide guidance on making use of subject headings. (Daniels &. Yakel, 2010)

2.2.4 Visibility

This section deals with the issue of making the archival holdings more visible, by allowing users to access them through Web search engines and union archival searches

2.2.4.1 Search Engines

We need to ask how to reach users who don't know about individual archival web sites; those who are not aware that the information they need is buried in the finding aid. It is the archivist's job to make information on their holdings accessible to new users, and a very important part of this is making the finding-aids searchable through search engines.

The problem is that currently archives which publish their descriptions on the Web usually found in the "deep" or "invisible" Web and the researcher has to know of the existence of the particular database in order to explore it further. This is because their contents are hidden from search engines and only accessible from search forms within their own web sites. (Hill, 2004; Kiesling, 2001). The reason for this is that Web crawlers usually access only the file header, which in this case holds information on the finding-aid itself, and not the description of the material. Access points that are of interest for users, such as, personal, corporate, and geographical names, topical subjects, and form and genre terms can occur within title statements and statements of responsibility, in paragraphs of text, and in segregated blocks. Those can be found at lower levels of hierarchy, e.g. item or file level, which are buried too deep within the document to trigger the web page searches (Kiesling, 2001).

The importance of allowing lower level descriptions to be made available to the search engines can be seen in the example of the portal providing on-line access to the descriptions of London and the M25 area called *AIM25*. This institution allowed collection level description to be crawled by search engines, which made usage of this service correspondingly high, with many users coming straight from search engines to the descriptions. Furthermore, a national gateway to archival collections held in UK called *Archives Hub*, has allowed Google's robots into the "news" section, which gives access to around 5 % of the Hub's descriptions in static web pages. This step has resulted in a huge growth of browse and index link searches: an increase of more than 500 % from

2002 to 2003. The results have shown that as many as 84 % of users came directly to these static pages from search engines. (Hill, 2004)

2.2.4.2 Union Search

The general user rarely knows which particular institution has the information or the document of his/her interest. For that purpose it is very useful to make the description of the holding of individual archives visible. One important way to achieve this is by contributing the description to some of the union catalogs. Also, as previously demonstrated, differences in institutional ways of description, representation of information and interface design, makes the user experience problematic. This means that providing union access to archives is desirable not only for end-user discovery, but also for accessing archival holdings.

However, some of the features of the archival description itself, and therefore EAD encoded archival description, hamper such attempts. This part of the literature review will with the problems within EAD DTD, that have proven difficult for the archivists themselves, as well as for the final product of their work, which is making information on their holdings visible to the end users.

One of the main reasons for developing the EAD standard is for the purpose of standardizing electronic representation of archival description, which makes it possible to provide union access to detailed archival descriptions and resources in repositories distributed throughout the world. This would allow a "global" user to locate archival materials and thus overcoming barriers that physical archives may impose to a researcher, such as their distant location or working hours. (Pitti & Duff, 2001). However, the literature suggests that this step is not so easy to achieve.

The main reason for this is that EAD working group purposely made permissive DTD. The key design principle states that EAD will accommodate both the creation of new finding aids and the conversion of existing (or legacy) data. (Encoded Archival Description

application guidelines, 1999). While the EAD DTD specifies the structure and syntax of finding aids, it does not mandate required elements, rarely dictates order and frequency in which elements occur and does not describe the form or nature of the content of any of those elements. For this reason many repositories have chosen to encode finding aids in a form very close to the legacy, following national or institutional guidelines that can vary from few to over 40 pages. (Shaw, 2001) As Shaw stated: "*Although this flexibility may have contributed to wide adoption of EAD, in the long run it hampers the very data exchange for which EAD was created.*" (Shaw, 2001, p. 117).

The main problem is that the use of encoding elements for the purpose of description varies greatly different institutions (Kim, 2002; **Prom, 2003)**. Furthermore, there are multiple ways of expressing information, as the DTD actually allows numerous ways of capturing similar information. Examples of this can be two solutions for identifying a series number: use of the <unitid> tag and use of the *id* attribute of the relevant component <c>. Another challenge to consistent machine and human processing is the extent to which elements are frequently available in multiple levels and places in a document, for example <unitdate> can occur within a number of places within the document. (Shaw, 2001)

Some individuals who have worked closely with EAD believe that its extreme flexibility undermines the goals of information exchange (Hoyer, Stephen & Pollock, 2001; Prom, 2002; Shaw, 2001). Prom (2003) suggests that the essential problem is that union catalogs cannot be built with records that do not share a minimal level of uniformity. (Prom, 2003).

This uniformity can be achieved by following some of the best practices guidelines proposed for encoding in EAD, such as guidelines that are considered to be most detailed and widely implemented standards currently available, e.g.:

-The *RLG Best Practices Guidelines for Encoded Archival Description* is published by the Research Libraries Group (RLG) issued in August 2002 (RLG EAD Advisory Group, 2002)

- *Library of Congress Encoded Archival Description Best Practices*, was published in 2008 (LC, 2008)
- *OAC Best Practices Guidelines for Encoded Archival Description* (OAC Working Group, 2005). (Carpenter & Park, 2009)

Another way is placing constraints on a data model for archival description, by assigning to the DTD at least required core descriptive elements. Such a consistently encoded, required core set of elements providing pointers to content in uniform ways, would provide data sets around which a generalized suite of data entry, retrieval and display tools could be built (Shaw, 2001). Furthermore, this would allow a federated search. This method has been utilized by service providers, such as archival portals that make use of the harvested metadata from other archives. One of such example is Archives Portal Europe- APENet project, (available at URL: http://www.apenet.eu/) which has developed a specific subset of EAD called APENet EAD, acting as a pivotal format for integrating data from all European archives that provide their institutional descriptions mapped to it, in order to allow union searches.

The Open Archives Initiate Protocol for Metadata Harvesting (OAI-PMH) presents one of the most used methods by which metadata regarding archives and manuscripts can be shared and made more interoperable with metadata from other sources. This is because interoperability issues around providing union access do not just finish with the reconciliation of EAD encoded data. As Prom (2003) states, many cultural heritage materials (e.g. personal papers, manuscript collections, organizational records, photographs, art pieces, maps, artifacts etc.) are managed by libraries, museums and other institutions that are not applying archival description, but describing them by one of the other existing metadata standards. For the purposes of incorporating these descriptions in union catalogs, and even providing cross domain search facilities, there are many interoperability possibilities . This thesis will investigate one of these, which is mapping EAD to a model that allows such an interoperability solution.

2.3 Chapter Summary

The archival public has dramatically increased by making archival description and its digital surrogates available on the Web. Some significant parts of this new public are users that do not possess the so-called "archival intelligence" that is necessary for conducting archival research. This skill is still as necessary in the on-line environment as it was in the physical, because standards, including the EAD standard, were made with the purpose of fully encapturing and copying the description on which physical finding aids were made. But this brings us back to the issue - is the physical finding aid really designed to facilitate access to the researcher or to the archivist who organizes the collection and acts as a mediator between it and the user? The content and format of finding aids, whether they are paper or web based, has not changed substantially during the last fifty years, however the archival user population has changed dramatically. (Gilliland-Swetland, 2001) Nevertheless, because of a lack of thorough user studies, we still don't know enough on their search behavior. Most of the studies conducted so far that surveyed in this literature review were conducted in controlled settings with predefined search tasks performed on specific institutional interface solutions. Those studies do give a certain consistent view on novice users and their behavior. The main message encountered in these studies can be best described by Shier (2004) who states each of the finding aids searched in her study, site structure, display, terminology, and navigational elements often worked against users' desire for immediacy, leading to much clicking and scrolling, and to a good deal of frustration.

In some of the researches, users voices can be heard, bringing this abstract contention to life:

frustration: "wasted a lot of time clicking in and out of the web site trying to find my subject matter in a sloppy and confusing manner." (Shier, 2004; p. 60)

helplessness: "I've got to look in about six places. Now I've got to figure out how to get through the system to the individual record. And it is a system." (Duff & Johnson, 2003, p. 91)

uncertainness: “I found it under Contents List by randomly clicking” (Daniels and Yakel, 2010, p. 556)

impatience: "I felt impatient about reading through all the text on the first page and wondered which link to follow. It seems to me that users just want the key to the finding aids as quickly as possible, so whatever gets them there fastest and with the fewest clicks is the best." (Prom, 2004, p. 25)

All of those cries of distress show that novice users do not cope very well with the current state of archival description offered.

Archives have a new duty to fulfill, and that is to allow new users to search for and discover their holdings. However, mounting finding aids, that is, providing networked access to them, does not make them accessible, discoverable, or useful.(Tibbo, 2003). In the theoretical framework developed, there are four main factors that currently hamper accessibility by general user:

1) archival terminology used in on-line presentation,

2) structure in which the presentation is organized and expressed (hierarchical structure),

3) access possibilities that support mainly browse and not search function,

4) unawareness of users of particular archival holdings, caused by the EAD structure in which a description can be found, which hampers Web crawlers from reaching information and union archival searches.

EAD as a format allows different display options. The studies done so far were conducted on one or more institutional interfaces that present EAD encoded finding aids, and from

them it can be concluded that a proficient archival researcher has no problem, and even supports such presentation and contextualization (Prom, 2002). Therefore, the research path I have chosen to follow is to make the already existing and widely accepted EAD model interoperable with a more general language (EDM) and investigate whether it may bring improvements for the non-proficient archival public. If it would bring improvements, this would mean that not one, but many archives could be mapped to a single pivot language, that would allow them to communicate with their general users on one side, while retaining the complexity and fullness of original data to communicate to the experienced users, on the other. The theoretical framework will serve as a guide to developing a general method for mapping from EAD to EDM, applying this method for the purpose of validation and criteria through which I will discuss possible improvements.

Both languages in question (EAD and EDM) are discussed in more detail in Appendix 2 and Appendix 3, respectively and it is highly recommended to read them before reading the Research Design, Chapter 3.

CHAPTER 3: Research Design

In this chapter the methodology chosen to conduct this research project will be discussed. The *aim* of this thesis is facilitating access to the archival records in on-line environment to wide general public. There are several ways of achieving this aim. The particular approach chosen in this thesis is to offer to the users a search language based on general concepts that should be more understood by the users, such as people, places, events, time and the like. These concepts are part of the Europeana Data Model (EDM), which is therefore used as a model of the underlying archival records *as far as the user is concerned*. The *objective* I am trying to reach is to investigate whether this model (EDM) would bring improvements to the general user population.

Since the archival records are not modeled according to EDM, the objective can be achieved by transforming the underlying archival records into EDM. There are two main ways of transforming the archival records into EDM:

1. to actually perform the transformation by migrating the archival records from their native model into EDM.
2. to simulate the transformation by mapping the actual data model of the underlying archival records into EDM.

I have chose the latter approach mainly because the former approach requires the ability of accessing the archival records themselves and re-writing them in the new format, and this is not always possible.

Before delving into the technical developments required for performing the mapping, there is however a question that must be answered. Is EDM good enough to play the role of "universal" access language, easily understood by the general public? In order to answer this question, I need therefore to investigate whether EDM model could bring some improvements to the on line archival access to general public**.**

Therefore, the research question that this thesis intends to answer is:

- **Would transforming EAD encoded archival descriptions in EDM bring improvements to on-line access for general archival public?**

In order to answer this question, it is considered as necessary to find an answer to three sub problems underlying the research question:

- ✓ *What do the members of general user population find problematic with current online finding aids, and what are their preferences for access?*

this sub problem was tackled in the literature review in Chapter 2, where the theoretical framework was built, outlining the issues thought to be important for this context.

- ✓ *How to develop general mapping method from EAD to EDM?*

The present chapter deals with this sub problem, proposing a possible solution to it.

- ✓ *How to apply this method to a real set of data?;*

Chapter 4 will deal with this issue, where the method developed in this chapter will be applied to two fonds from the archive of *Accademia Nazionale di Santa Cecilia*, a music academy in Rome.

3.1 Methodology of the Research

In order to try to give an answer to the research question stated above, it is necessary to achieve the conversion from EAD to EDM. Therefore, the general methodology applied in this research is of prescriptive nature. The methodology applied is inspired by the Design Science, which aims at "producing and applying knowledge of tasks and situations in order to create effective artifacts" (March and Smith, 1995, p. 253). The artifact in this case will be a *method* to map EAD to EDM that would allow the "conceptual" mapping of the two schemas in the first place, and at the same time providing the basis for developing a tool for an automatic mapping. The axiology of

Design Science research stresses problem solving and an effective way of accomplishing an end result. (Vaishnavi & Kuechler Jr, 2007) The output of the particular methodology applied in this case will be to derive a *method* for mapping the EAD schema to the EDM model. *Method* is defined as a "set of steps (an algorithm or a guideline) used to perform a task. Methods are goal directed plans for manipulating constructs so that the solution statement model is realized." (Vaishnavi & Kuechler Jr, 2007). The criteria for making decisions and choices in developing this method will be guided by the objective of conveying as much as possible of the information contained in the original EAD data to the EDM. The development of the methodology will also be based on the literature review and the suggestions from archival experts and researches, in order to try to solve some of the problems that a general user is facing when accessing on-line archival data.

In order to investigate, on the theoretical level, what possibilities the EDM model could bring to EAD encoded data, the methodology applied in this research project was considered as the most appropriate, as it allows the possibility to compare the original and newly modelled data and discuss the results. Furthermore, it was considered necessary to perform a structural analysis of the two models in question to gain a better understanding of them. These analysis can be found in Appendices 2 and 3, for the convenience of the reader (as well as for the author).

3.2 Developing a Method for Transforming Archival Data Encoded in EAD to EDM

The EAD data has a hierarchical structure with descriptions associated with the nodes of the hierarchy. For this reason it is convenient to divide the general problem of mapping EAD into EDM into two parts: the structural mapping, i.e. the transformation of an EAD hierarchy into an equivalent RDF graph; and the metadata mapping, that is the transformation of each description found in an EAD record into an equivalent EDM description.

It is important to notice that the two parts are related, as in EAD the metadata records of a node implicitly also apply to their sub nodes. This aspect will be considered in due course.

3.2.1 First part: Structural Mapping and Aggregation Building Introduction

This part explores in which way to transform the archival description structure (encoded in EAD), which is **"tree",** into the structure of EDM, which is **graph based** model, while at the same time keeping the original information found in archival data. The steps of the transformation are the following:

1. transforming EAD tree node C into an EDM Aggregation A
2. associating an OAI-ORE Proxy P to the Aggregation A, by means of the OAI-ORE property ***isProxyIn***;
3. using the Proxy P as a representative of the real-world entity that node C is about, by means of the OAI-ORE;
4. using the DC property ***hasPart to*** relate the proxy P with the proxies defined for the children of node C in the EAD tree. In this way, the EAD tree is represented by the tree induced by the ***hasPart*** property;
5. retaining the order of the sibling nodes of C by means of the property ***dcterms:IsNextInSequence.***

In the rest of this section I am going to illustrate in detail each one of these steps.

3.2.2 Step : Transforming EAD nodes to EDM Aggregations

As already stated, the EAD data are hierarchically arranged, usually comprising several levels of descriptions, as schematically shown in Figure 14.

In the Figure 14, it is demonstrated that the description of the archive itself (<archdesc>) is considered to be the *root* node. The first level component units (in this case <c1>) are

its direct *children* nodes and are considered to be *subtrees* of the root node. Going down into the hierarchy, the next component level (<c2>) is considered to be a *subtree* of level <c1> and so on. The nodes that do not have any sublevels are considered to be *leafs* of the tree. The *sibling nodes* are those having the same parent element.

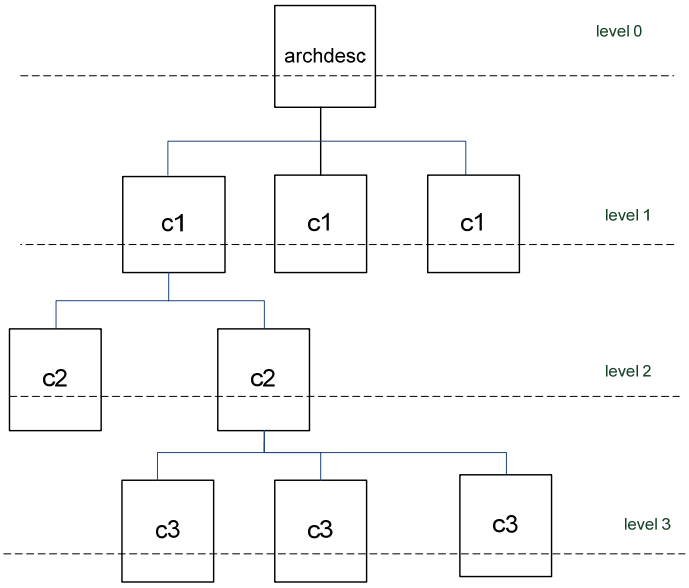

Figure 14: Tree structure of EAD file, without <eadheader>

On the other hand, the structure of EDM data is expressed by OAI-ORE through *Aggregations*. An Aggregation can be described as a set of related resources (e.g. the object itself, its digital representations, and descriptive metadata about both) that are conceptually grouped together, so this set can be treated as a single resource.

In order to preserve the original structure we need several Aggregations to model the complete EAD data. Therefore the recommendation for this part of the mapping is to **create an Aggregation for each node found in the hierarchy (with some exceptions).**

The first exception is the node corresponding to the <eadheader> element in the EAD data, which is usually the very root of the whole description. The reason is that this element contains the *Description of the Finding Aid* itself, and once that the EAD data is remodeled in EDM, the Finding Aid does not exist anymore, and therefore there is no

need to create an Aggregation for this node. This decision was based on the outcome of a recent meeting between the EAD and EDM communities (see: EDM Archives meeting, 2010), where it was also suggested that if some of the information contained in <eadheader> would be relevant for the end user (for discovery or display purposes), then it should be saved by attaching it to some other appropriate Aggregation.

As stated before, structural mapping is going to deal only with notions that present and describe the archive and its material. Consequently, the highest hierarchical node is considered to be description of the entire archive (<archdesc>), followed by the levels of archival description (<c levels>). Therefore, these nodes will be presented as **Aggregations (**class name ***ore:Aggregation*****).** In order to declare what newly created Aggregations consist of, relations between different Aggregation needs to be declared using the property ***ore:Aggregates*****.**

In some cases a node does not require to create an Aggregation on its own. The decision whether to create an Aggregation for a node or not belongs to the person performing the mapping, and should be based on the information contained in the node and how relevant is the information carried by that node. Omitting the creation of an Aggregation for some nodes would, in some way, "flatten" the original hierarchy, but in any case the information contained in the node can be maintained, by transferring it to its children nodes. Some of the remarks in section 2.2.2 clearly indicate that the hierarchical structure of archive descriptions is one of the main causes of problem for novice user, and therefore omitting some nodes may be even desirable for the purpose of on-line access.

The figure below (Figure 15) shows how an EAD hierarchical structure can be transformed into a EDM graph.

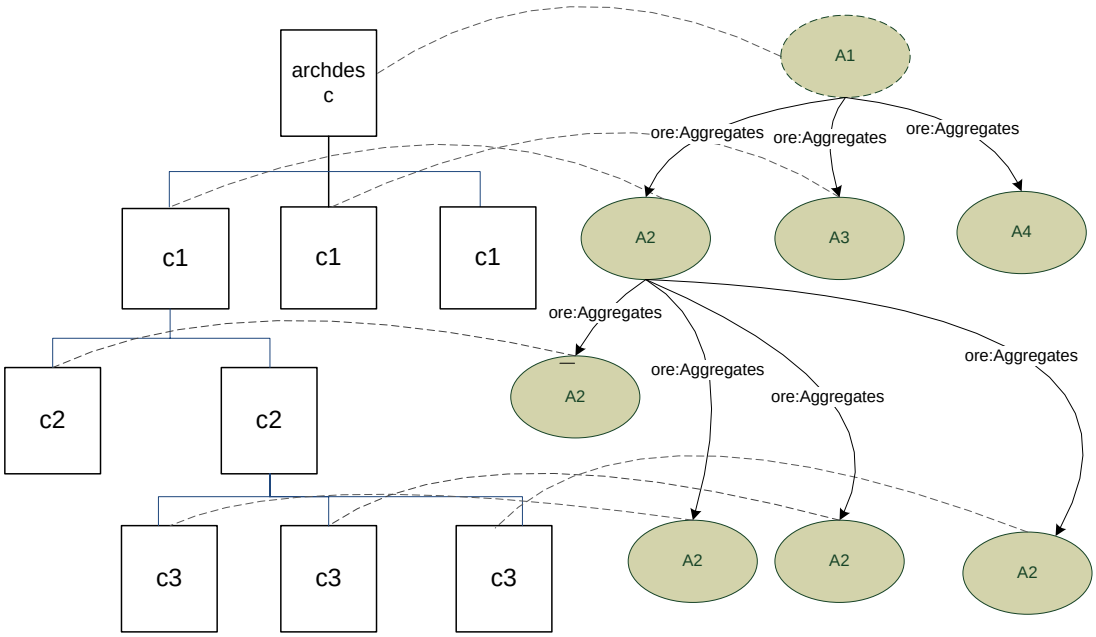

Figure 15: Creating aggregations for the hierarchical nodes

3.2.3 Step: Associating OAI-ORE Proxy to an Aggregation

For the purpose of carrying over the information that can be found in each node of the archival description, each Aggregation will be assigned with the Proxy (***ore:Proxy*).** Each proxy will represent the object associated with the Aggregation as viewed by one particular archive. Europeana uses proxies as place-holders for Cultural Heritage Objects within Aggregations for the purpose of making assertions about the object while distinguishing the provenance of these assertions. In summary, the proxy mechanism allows declaring different statements (possibly in conflict with one another) about the same object. (Doerr et al., 2010.)

The proxy is the entry point for metadata search on the objects provided, therefore all the metadata of a particular object is to be attached to the proxy. This metadata is going to be mapped to EDM properties, as will be described in one of the next sections.

3.2.4 Step : Associating Proxy and Aggregation to the Real Life Entities

As one Aggregation will carry description about the actual physical thing (e.g. the painting of Mona Lisa) and its eventual surrogate (e.g. the JPEG representation of the painting of Mona Lisa), it is desirable to make a distinction between the two. This may be done by assigning URIs to both the physical object (class ***ens:PhisicalThing)*** and its surrogate (e.g. class ***ens:WebResource***) and connecting them to the Aggregation and its Proxy, as shown in the Figure 16:

.

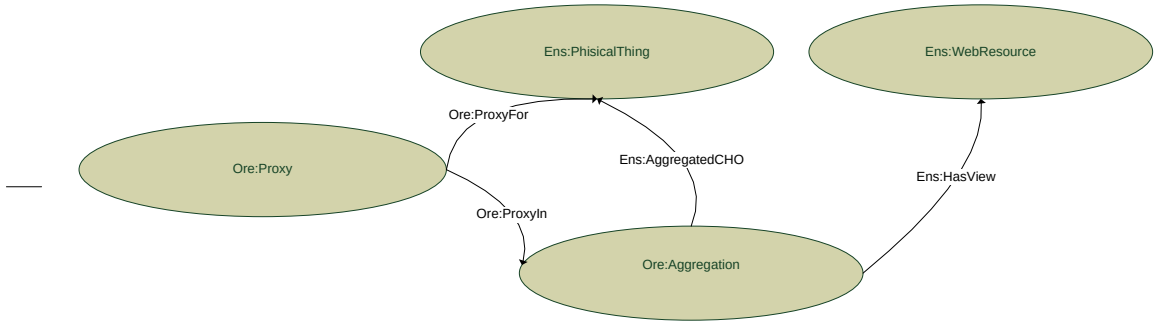

Figure 16: Connecting Aggregation and Proxy to the Physical Thing and Web Resource

3.2.5 Step : Representing Hierarchies in EDM

The hierarchical structure in EDM can be expressed by relating one resource which is included (either physically or logically) in the described one, by using the ***dcterms:hasPart*** property. Such relationship however is not established between the *Aggregations* themselves (*Aggregation* relationships are shown in Figure 15 and 16) but between their *Proxies*. The *Proxies* are related because they are the carriers of the metadata contained in the nodes. As previously described, in archival description the metadata of the higher level is inherited by its lower ones. By relating *Proxies* of different level, we achieve the same effect, which is explicitly stating the a higher level *Proxy* has

as its part the lower level one. This step results in retaining the "information hierarchy" from the original EAD data. (How this appears is shown in Figure 17 in Chapter 4).

3.2.6 Step: Retaining the Order of the Sibling Nodes

Archival material of the same level is very often ordered sequentially. For example, in the Audio Video fond that will be used to validate the methodology, at the "file" level we have the description of the recording of one concert, and at the "item" level we have the description of the different tracks, ordered as "Track 1", "Track 2" and so on. In order to keep this sequence when modeling the data in EDM, the property ***ens:isNextInSequence*** is used.

The definition of this property is : "***ens:isNextInSequence*** relates two resources R and S that are ordered parts of the same resource A, and such that R comes immediately after S in the order created by their being parts of A."(Definition of the Europeana Data Model elements, p. 22)

However, in order to retain such sequence in EDM, the ordering needs to be explicitly described in the archive metadata e.g. by using <num> element. When the information is implicit (e.g the ordering was embedded in the title, like "Concert Season 1954", followed by "Concert Season 1955") the ordering can not be retained in EDM.

3.3 Second part: Metadata Mapping

The first part has described the structural mapping of the archival metadata to the EDM, made in order to represent in EDM the hierarchical structure of EAD encoded finding aids, but also building Aggregations for Cultural Heritage Objects of one archive. The following part will deal with the mapping of the metadata that describes those objects.

"Metadata mapping is the process of identifying equivalent or nearly equivalent metadata elements or groups of metadata elements within different metadata schemas, carried out in order to facilitate semantic interoperability. Semantic interoperability is the ability to search seamlessly for digital information across heterogeneous distributed databases as if they were all part of the same virtual repository." (Baca, 2003, p. 49)
Furthermore, Baca emphasizes that "intellectual process of metadata mapping must be done by knowledgeable human beings familiar with both the intellectual content of the particular information resources and the various metadata schemas being mapped." (Baca, 2003; p.51). Following this "intellectual mapping" a developer can implement an algorithm in order to (partially) automate the translation from schema A to schema B.

For the purpose of developing this general method for mapping EAD standard elements to EDM model ones, in the first place the structural overview for both models was made (it can be found in EAD and EDM annexes) and afterwards consulted, together with the precise element definitions that are available for the both models .

The mapping to EDM properties is also a good opportunity to examine the EDM data, and make corrections to possible inconsistencies or errors made while creating them. As it has been identified by Dushay and Hillmann (2003) and Hillmann, Dushay and Philips (2004) there are four categories of problems that may limit the usefulness of metadata: (1) Missing data, e.g. elements not present, (2) Incorrect data, e.g. values not conforming to proper usage, (3) Confusing data, eg. embedded html tags, improper separation of multiple elements, etc., and (4) Insufficient data e.g. no indication of controlled vocabularies, formats, etc. The mapping, which requires thorough examination of all the metadata elements, is an opportunity to make changes to the original records, that would allow better access on the part of the users both trough institutional application (using original EAD data) and for the purpose of remodeling data in EDM. Some recommendations specifically regarding archival users could be found in the literature review, out of which maybe the most important one was to provide description at the lower hierarchical level (item level). Other recommendations to better exploit the potential of EDM will be made in a subsequent section.

3.3.1. Attaching Metadata to Proxies and Mapping to EDM

The first step is attaching the metadata describing the object to its appropriate Aggregation. To be more precise, this metadata will not be directly attached to the Aggregation, but to its Proxy, as they are the users' entry points for search. (EDM Archives meeting, 2010) Metadata attached to its corresponding Proxy is to be further mapped to appropriate EDM properties. To be more precise, EAD *elements* and their possible *attributes*, should be mapped to corresponding EDM properties. It is worth mentioning that the representation of EDM data is with RDF, which represents attributes through the notion of "properties".

3.3.2 Choosing the EDM Property for Mapping

To find in EDM a property equivalent (or as close as possible) to a source element, the EDM element specification should be consulted in order to see the definitions, constraints and examples of usage for EDM all classes and properties. When mapping to EDM properties, one should choose those carrying as much as possible semantic similarity to the elements or attributes of the original metadata schema. The effort should be made to stay as specific as possible in order to retain as much original information as possible.

EDM offers a range of properties, which are mostly defined in Dublin Core and Europeana namespaces, and to which more specialized ones can be attached and declared as subproperties. This should be done only when it is considered that elements from the EAD schema carries important information but in EDM does not exist a specific enough property to map it to. In this case, after finding first the correspondent generic EDM property to which map one EAD element or attribute, a specific subproperty can be declared, that will carry the information specific to the EAD model or an individual archival schema. For example, in EDM, the general property to express any kind of agent responsible for making contributions to the resource is ***dc:Contributor***. The property *contributor* is obviously a very wide notion and for description and discovery of some particular resource it may important to distinguish between the "contributor" W. A.

Mozart, as the *composer* and the "contributor" Nicolo Paganini as the *performer* of the resource described. It is then possible to create two new properties e.g. one called ***Composer*** and one called ***Performer*** and declare them to be the sub-property of ***dc:Contributor***. In such case, the EDM/DC properties can be seen as a kind of containers or anchors that would allow for

> -general searches among the broad semantics of the containers
>
> -specific searches among the specialized semantics of the sub-properties. (Hennicke, 2010)

Examples of this process can be seen in Chapter 4, where this methodology is validated with the actual data.

3.3.3 EAD Elements NOT Considered for Mapping

The aim of the mapping to EDM format is to facilitate access to archival holdings on the part of the general public. From this point of view, not all of the information that can be found in EAD metadata records may be relevant for actual discovery or presentation purpose. Gilliland-Swetland (2000), made a distinction between *administrative*, *descriptive*, *preservation*, *use* and *technical metadata*. The last group, for example, may be not too relevant for access, as technical metadata may capture information about e.g. hardware and software used for digitization, compression ratios, scaling routines, and similar. Out of this type of metadata only those elements that are considered to carry important information for the user, and not only for the internal institutional purpose, should be mapped. Properties in EDM that support technical metadata are, for example ***dcterms:ConformsTo*** (for the information on standard), ***dcterms:extent*** (for size or duration information), ***dc:Format*** etc.

Additionally, EAD has a number of elements that do not carry specific semantic information, but their purpose is to group other elements. They are also called "wrapper" elements (eg., <did>, <descgrp>, <dsc>), which should not be considered in the mapping process. Furthermore, some elements and attributes that are used for formatting purposes

(eg. <p> for paragraph, <list> and other) also should not be mapped, but translated as XML-literals and if necessary rendered appropriately.

3.3.4 Metadata Mapping Methodology

EAD data (represented in XML files) are usually very complex and are using composite elements to encode information. Composite elements are those that contain other elements and/or attributes. Therefore, when mapping, a single element is not mapped separately from its ancestors, but the ***path*** in which it can be found will determine the EDM property to which it is mapped to. For example, the element <num> when found in path **c/did/controlaccess/name/num** is mapped to ***dc:Title***, as it carries the name of the track title, but if found in path: **c/did/unititle/bibref/num** it is mapped to **dc:Description,** as it carries the value of a ordering number of track (example from the Ethnomusicology fond).

The core idea behind converting EAD data into EDM is that every **complex** element maps to a resource, i.e. a node (often a blank node) and every atomic attribute maps to an attribute of this node. Such a translation would give a valid EDM graph, which is much easier to access for further processing. However, for further processing, the creation of blank nodes should be avoided as much as possible, as they blur the information and make query answering difficult. (RDF primer, 2004) For this reason these blank nodes should be replaced by a proper URI, to be declared as an instance of the EDM class (see EDM classes in Fig. 7)

3.3.5 Event-centric Modeling of Data

As it can be seen in Appendix 3, in addition to object-centric representation of the described resources, the EDM model also allows an *event-centric* modeling of data. Therefore if the archival data holds information about a particular event (usually encoded in <event> and paired with <date> element) it is possible to represent this event by creating an instance of ***ens:Event*** class. This mechanism is highly desirable, as it answers

to the preferences expressed by genealogist and institutional users (see sections: 2.2.3.1.1 and 2.2.3.1.2). The created events can be linked to all the important actors related to them by properties ***ens:hasMet*** and **ens:WasPresentAt**, as well as specifying time and place of the event by properties **ens:occurredAt** and **ens:happenedAt** (See Appendix 3, Figure 13).

3.4 Practical Issues

The previous text was regarding metadata mapping methodology on the schema level. However, it is important also to address the data value level that should be taken in consideration when modeling the data in EDM.

3.4.1 Thesauri

The use a controlled vocabulary (also in finding aids) would be very desirable, as it would improve the recall of searches and would help in generating search terms. However, this is not an usual practice for the archives to index their objects with a thesaurus. In the EDM world, in order to achieve semantic interoperability with other data encoded in EDM, it is recommended that possible thesauruses used be translated in SKOS (this process is also known as "SKOSification") (see Appendix 3, Figure 6). Currently, the potential for applying SKOS in the archival area is limited, as it can only provide very limited semantics given that the original indexing sources do not contain structured semantics. (Olensky, 2010) For these reasons, it is suggested to use controlled vocabularies and translate them in SKOS.

3.4.2 URIs

Another issue is providing identifiers to the objects of importance in collection. For the purpose of translating the data to EDM use of Uniformed Resource Identifiers (URIs) is advisable in order to fully exploit the possibilities this model is offering. URI is a string

of characters used to identify a name or a resource on the Internet. “URIs provide a global naming scheme which allows immediate interoperability between any data sets expressed in RDF”. (Styles, Ayers, & Shabir, 2008) Having URIs would result in richer use of data such as semantic search (Guha, McCool, & Miller, 2003) for which reason institution may choose to generate them. However, this is the issue of institutional policy. Otherwise, they will be created at the time of normalization of data by Europeana.

For the creation of a URI, a naming schema should be defined, and there are many different ways to do it. One possibility is based on the use of the (literal) values that can be found in data, which can be used to construct a unique identifier (a URI) for:

1. Objects that are unique and specific to the collection, which is the typical case with cultural heritage objects appearing in a collection. The collection usually provides a unique object identifier which is a perfect basis for defining a URI.
2. Things that appear in multiple collections, such as *persons, institutions, styles*, etc. are candidates to be linked to background knowledge from vocabularies. Many of these resources have proper names and it is desirable to create a resource based on this proper name, possibly augmented with disambiguation information such as from a hierarchy. e.g. Amsterdam-Netherlands. (Milestone, 1.3.1, n.d) Basically, a URI should be made for everything that is in the core of institutional interest.(Minutes, 2010) Another method for generating URIs can be seen in the paper by Styles, Ayers and Shabir. (2010).

3.4.3 Usage of Already Available URIs for Indentifying Institutional Resources

As stated in section 2.2.3.1.1, it would be of great usefulness for the single biggest group of archival users, which are the genealogists, to have the possibility to locate documents by the name of place name, regardless of possible changes that occurred to it over time. The same consideration applies to the personal names and the disambiguation between persons with the same names. One of the goals for which the EDM model was created

was also to connect to the Linked Open Data cloud, in order to reuse, as much as possible, already available Web resources that would allow an enrichment of data, which in turn would answer to some important user needs. In conclusion, it is very convenient to use URIs that already exist in the Linked Open Data cloud to indentify resources in institutional metadata. An example of one such international effort for the authority files of personal and corporate names is The Virtual International Authority File (VIAF) (http://viaf.org/), implemented and hosted by OCLC. As stated on its Web site, is a joint project of several national libraries plus selected regional and trans-national library agencies and its goal is to lower the cost and increase the utility of library authority files by matching and linking widely-used authority files and making that information available on the Web. However, still it is not clear which exact data base should be used for exploiting URIs and how to deal with the cases where such authority files can not be found for every instance considered to be of special interest for the institution.

3.5 Chapter Summary

In this chapter I have shown the research design of this study and developed a general method for mapping EAD schema to EDM model. This method consists of two main parts: first the structural mapping, and then the metadata mapping, which are further divided into different steps. In the following chapter the method developed is going to be applied to real life data for the purpose of validation and discussion.

CHAPTER 4: Validation and Discussion

4.1 Validation Introduction

This part of the thesis will discuss the validation of the previously developed method for mapping EAD model to EDM schema, by applying it to a real life example of EAD encoded data. This step is considered necessary because of the very nature of the EAD standard. EAD was purposely defined to be a very permissive model in order to accommodate all the different archival descriptive practices. (Shaw, 2001) As shown in the introduction and literature review, this standard consists of as much as 146 elements, of which only few are obligatory, and many attributes that specify the further use of these elements. To completely validate this methodology by mapping the entire EAD schema to EDM would require a very long time and a study of much greater extent. For the purposes of this thesis, which is a master level and time limited one, I have performed the validation using two fonds provided by the archive of the *Accademia di Santa Cecilia*, which are using just a subset of the complete EAD schema. This step was considered a necessary part of the methodology developed in this research, as it would allow validating the method developed in Chapter 3 and finally commenting and analyzing the changes that took place through this modeling exercise.

The data chosen for the validation and analyzing part is coming from the Multimedia Archive (Bibliomediateca) of *Accademia Nazionale di Santa Cecilia* (in further text called ANSC). ANSC is an internationally renowned musical academy located in Rome, Italy and one of the oldest musical institutions in the world. The entire patrimony of this institution is about 120,000 volumes and publications, mainly scores, monographs and periodicals about music.(Bibliomediateca, n.d) Two fonds from this archive were mapped to the EDM model for the purpose of validating the method and analyzing the process and possible results: *Ethnomusicology Fond* and *Audio Video Fond.*

Ethnomusicology Fond (Fondo Etnomusicologia) consists of valuable recordings of Italian oral music traditions, assembled since 1948 thanks to the field research of the *National Center for Studies in Folk Music (Centro Nazionale di Studi sulla Musica Popolare)* founded in collaboration with the RAI, the Italian radio and television national company. This fond is the result of the work of Giorgio Nataletti, Diego Carpitella, Ernesto De Martino, Alan Lomax and other scholars who undertook the research and study of this particular repertory. Further details and on-line access to this fond is available in English at the URL:
http://bibliomediateca.santacecilia.it/bibliomediateca/cms.view?munu_str=0_1_0_5&numDoc=277&l=EN

Audio Video Fond (AudioVideoteca) contains sound heritage built up during the 20th century through numerous donations, important collections of recordings on vinyl, tapes and in digital formats. In this fond, the performances of legendary conductors of the past (Toscanini, Furtwängler, De Sabata, Molinari, Karajan) are kept together with those of the finest conductors and performers of today, and is entirely available for public. Further details and on-line access to this fond is available in English at the URL:
http://bibliomediateca.santacecilia.it/bibliomediateca/cms.view?munu_str=0_1_0_3&numDoc=275&l=EN.

The main reason for choosing this archive and its fonds (in addition of course to the fact that their finding aids are EAD encoded) was that ANSC is a partner in the ASSET project, as well as ISTI-CNR, the Institute where I have conducted my research. One of the main objective of ASSETS (see: http://www.assets4europeana.eu/ for complete details) is to make more digital items available on Europeana by involving content providers across different cultural environments. The interest of ANSC to be a content provider to Europeana (and therefore to make its content available in EDM) and the good relationship between ANSC and ISTI-CNR (both being partners in ASSETS) have allowed not only the possibility of accessing and having available all the archival material, but also have allowed an easy communication with the ANSC staff and encoders, in order to clarify encoding choices and the semantics that particular elements

and attributes are carrying in the ANSC archive. Finally, this archive was fitting very well with my "past" background, having studied and played flute for about ten years (before deciding to go to a LIS school rather that to a music academy), acquiring a good knowledge in musical matters and terminology.

Although, it would have been desirable that all the fonds contained in this archive were covered by this validation, this was not possible due to the limited time in which this thesis had to be conducted. For this reason, two representative fonds were chosen.

4.1.1 Pre-processing Steps

The descriptions of the two fonds chosen was made available as two separate EAD XML files, and each fond was processed separately. The DTD used for creating those files was the official EAD 2002 DTD (available at URL: http://www.loc.gov/ead/ead2002a.html). However, as already stated, the EAD schema is very permissive and general in order to accommodate many different archival practices, and therefore not useful for the purpose of the detailed analysis of one particular institutional subset. Following the methodology developed in Chapter 3, stating that hierarchical nodes become Aggregations to which metadata is attached, the metadata found under different levels of the hierarchical nodes have been separated, in order to examine possible differences in the paths of different levels. It has to be noted that in the ANSC data, levels are encoded by the value of the attribute "level" in the <c> data element:
<c level= "fond|series|subseries|recordgrp|file|item">

The separation of the different levels found in the description of the fonds was performed by using ad hoc software developed at ISTI-CNR. For each extracted level a separate XML file was created, and its DTD schema was derived (by means of *Altavista XML Spy* software), in order to capture all possible combinations of element occurrences that could be found in that level, so that the mapping of the nodes of a given level would cover all the possible elements.

The DTD schemas extracted from each level have been analyzed by means of two XML editors: *XML copy editor* and *Altavista XML spy,* and their element have been recorded in the mapping tables described in the next section. This mapping methodology is based on Path-Oriented Approach. (Theodoridou and Doerr, 2001) It is important to underline that only in those cases where different paths of elements carrying different information were noted, the levels were separated in the mapping tables, otherwise they were mapped as a single level.

4.1.2 Mapping Tables Description

Most of the work for validation part of the thesis is summarized in the metadata mapping tables (see Appendices 5 and 6). In the far left column (a) there is the path in which the EAD elements were encountered in the original file. The next column (b) describes semantics of these elements, based on which the most appropriate EDM counterpart was chosen (see column d). At this point, it is worth to note again the differences of institutional practices when it comes to describing resources, which is the reason why usually the original data has been used to explain the semantics of elements, rather than the EAD 2002 Tag Library. The third column (c) holds the default values of the attributes found, while the fourth column (d) holds the EDM properties to which the EAD ones have been mapped. The last column (e) holds the RDF objects created for the composite elements (whether they are blank nodes – BN, or instances of EDM classes), and possible notes on how to deal with them. The numbering of the created RDF objects (blank nodes and class instances) was made only to facilitate the mapping activity. The table does not include the information element <processinfo> as in the ANSC data it contains information on data entry or changes in the original files, which has not been considered as relevant, as shown in the Chapter 3. Other elements that have been omitted in the mapping can be found at the end of the tables, and are basically those used in EAD for formatting purposes. The elements not considered in the mapping are labeled with the "/" character, and those elements that have the same mapping in different levels are labeled with the "~" character. Other characters used in column (d) are described in the next sections.

4.1.3 General Notes on ANSC Data

The archival data provided by ANSC differs in many aspects from the "traditional" EAD data usually found in other archives. The ANSC data was not containing the data elements <ead> and <eadheader>, which in any case would have been left out from the mapping. Another element missing was <archdesc>, which usually is the highest hierarchical node. In ANSC, the highest node is always the data element <c>, with the value of the "level" attribute equal to "fond":
<c level= "fond">,
which contains the entire fond in question.

Another distinctive feature of the ANSC data was a rich description in the lower hierarchical levels, and much less in the upper levels, contrary to the usual archive practice. As noted in the introduction, usual archival practice is to describe the upper hierarchical levels more fully, since this description will be inherited by the lower ones. It is not a frequent case where rich descriptions are made for the lower levels (especially at the file and item level), although this is highly recommended. (Gilliland-Swetland, 2001) From this point of view the ANSC data is closer to user needs than that usually found in other archives.

Another particularity encountered in the ANSC data is the use of the attribute "audience". Usually this attribute is used to control whether the information contained in the element should be available to all viewers (in which case the value is "external") or only to repository staff (in which case the value is "internal"). If the attribute is used in this way, than element in which it can be found should not be ingested by Europeana, which is clearly target to external users, and hence not mapped to EDM. However, in the case of ANSC data the value "internal" for the "audience" attribute was used to explicitly state that the actual contents (the object described) should not be published, at least not at certain resolutions and formats (e.g. the MP3 audio) but still the general public should be able to locate it through the metadata assigned to it. For this reason all the <c> data elements have been mapped to EDM.

The ANSC data contain also a section holding data not originating in EAD. This data was originated from the MAG standard (Metadati Amministrativi Gestionali) an Italian standard maintained by ICCU, the (Central Institute for the Union Cataloguing). (Standard MAG, 2011) The section encoded by means of this standard can be found under the element <metadigit> (see: ETN table element 57 and AV table element 116) and it is intended to keep the technical metadata of a digitized version of the object described.

One sample of ANSC XML data, i.e. "branch" of the Ethnomusicology fond is shown in Appendix 4. This fond has only three levels and is therefore used for a clearer demonstration of the mapping. How the sample xml "branch" looks after transforming in EDM can be seen in the graph shown as the Figure 19. The other fond, Audio Video, is much richer in terms of metadata it contains as well as the levels (four were encountered), therefore the sample was not included because of its length, but the mapping steps applied are the same for the two fonds, and are described in the next sections.

4.2 Validation Examples

The examples in the next sections apply to both fonds. Where I want to stress the work done on Ethnomusicology fond I will address it with the abbreviation **ETN table (fond)-(see: Appendix 5)** while the work done on Audio Video fond will be addressed with the abbreviation **AV table (fond) (see: Appendix 6)**.

4.2.1 Structural Mapping

The structural mapping and the steps that should be taken in this part of the work are described in chapter 3.2.1. The first step has been the creation of Aggregations for the hierarchical nodes in the fonds. As already mentioned, upper hierarchical levels in both fonds have scarce metadata in them, in my opinion insufficient to model them as separate "objects". Therefore I have decided NOT to make an ***ore:Aggregation*** for them (see 3.2.2

paragraph 7) but rather to attach the information they are carrying to the next level. In order to do this I have made new sub-classes of ***ens:NonInformationResource*** called ***ens:ArchivalFond*** (for the highest level node in both ETN and AV) and ***ens:ArchivalSeries/ ArchivalSubfonds*** (for the next level in AV). Those new classes became the range of property ***dcterms:isPartOf***. For the remaining levels (file and item) I have created instances of classes ***ore:Aggreagations*** and ***ore:Proxies*** and linked them in order to copy the hierarchical structure, as described in sections: 3.2.2 and 3.2.5. This part of the mapping can be represented in the Figure 17

:

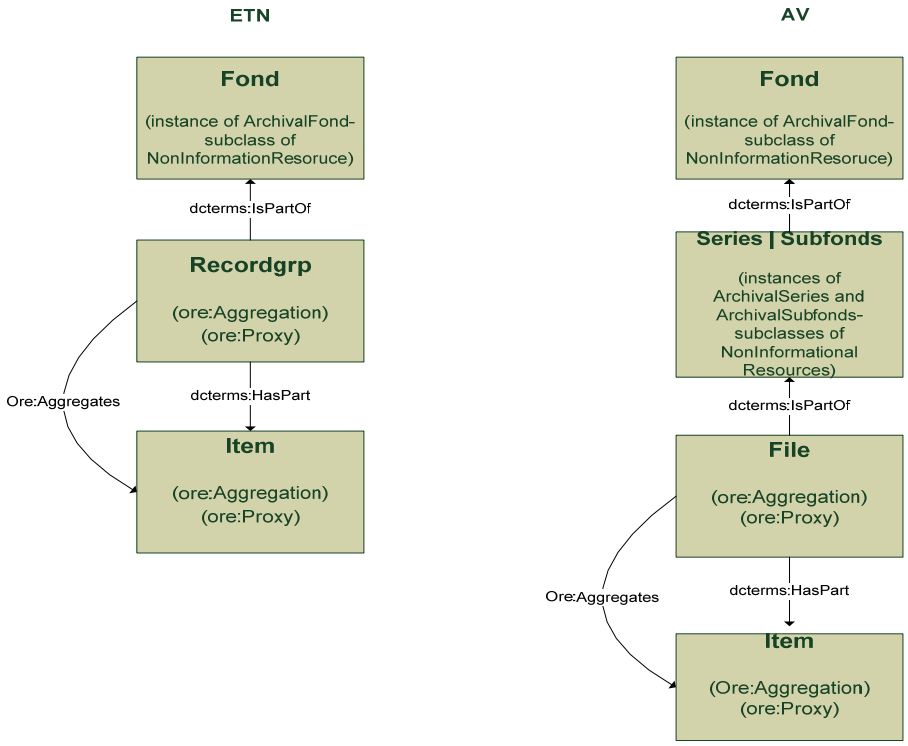

Figure 17: Structural Mapping of Ethnomusicology and Audio Video Fonds

The other aspect of structural mapping is to maintain order amongst elements at the same level as established in original documents (see section 3.2.6). This is done by linking elements at the same level that have explicit encoding of ordering numbers in EAD with the ***ens:NextInSequence*** property. For example, the element <num> of the ETN fond at

the level **record group** holds the ordering number of the collection described, and on the **item** level it holds the ordering number of the track in this collection. (see element 26 in ETN Table). In AV fond a possible sequence can be made, at the file level, by using the <sequence_number> element (see element 160 in AV table.). At this level the <num> elements encode the Title and not the number, and at item level a possible sequence can be made by using the <num> element (see element 170 in AV table).

4.2.2 Metadata Mapping

In the next step the metadata describing the nodes of a given level will be attached to the **ore:Proxies** created for representing the node. All the metadata that can be found under one level of nodes is separated and can be seen in mapping tables.

Some elements have been identified as non-eligible for mapping (see section 3.3.3.). They are mainly "wrapper" elements (elements that do not carry specific information, whose main purpose is to group together the elements within it) and the technical metadata. For those elements, in the mapping tables, in column (d), instead of having an EDM property the character "/" was used to signal that no mapping was performed. Also, elements used for formatting purpose were not included in column (a), as they should be rendered as XML literals. They can be found at the end of the tables (<imprint>, <emph>, <chronlist>, <list><cronitem>, <item>, <head>, <p>).

4.2.3 Mapping Composite Elements

In the section 3.3.4, it was described how the composite elements should be dealt with. Column (e) of mapping tables shows how I have applied this approach in my validation The graph below (Figure 18) is an example of mapping of composite elements, which includes instances of classes ***ens:Place*** and ***ens:Agent*** (see elements 1-10 of ETN mapping table). This is shown on the example of the instance made for the newly created class **ens:ArchivalFond**.

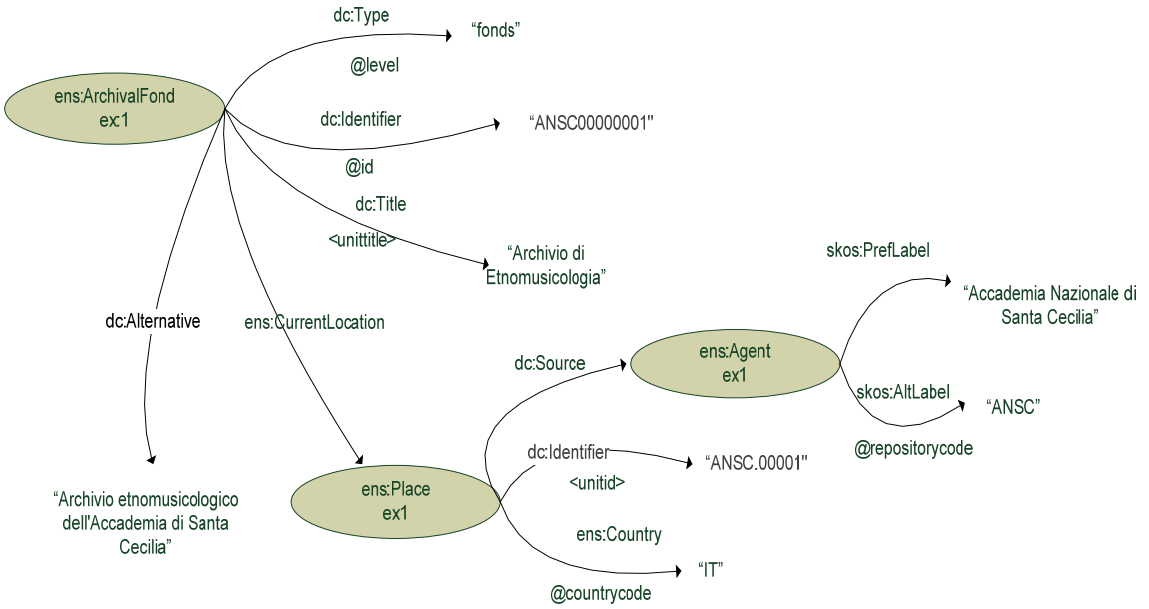

Figure 18: Examples of composite elements mapping

4.2.4 Issue: Creating New Classes and Properties and Alternative Solutions

As it can be seen from the EDM property taxonomy (see Figure 8) the properties offered by EDM are very general and sometimes for institutional purpose it is preferable to keep the specific semantics of the elements found in institutional schema. In order to do so, new properties and classes can be declared as specializations of the EDM properties and classes. In this section I will show some of those cases, including cases where existing properties can be re-used.

In both ETN and AV fonds, but also in many other archival descriptions, a notion of Abstract (<abstract>) can be found. This information should clearly be retained, but there is no such property in EDM. In this case, since in Dublin Core there is a ***dc:abstract*** property, defined as a refinement of ***dc:Description***, it was decided to use this property instead of defining a new one.

Another approach deals with the **different attribute values** that can be found in data. In XML attribute specifies the element and sometimes different values of the attribute call for different values of the element. In the cases where those different values are considered to be information I wanted to save, for example for display purpose, but are not considered as the information through which the user may query the data, SKOS naming features were utilized.

For example in the AV mapping table (see element 29, with a "*" in column d) the element <name> was used and carried different values, depending on the occurrence and values of its attributes. The attribute @role had different values such as: "organico", "forma" etc. If <name> occurred together with attribute @role, then the following mapping was made: <name> was mapped to ***dc:Description***, which had the new value-***skos:Concept***, which in turn had all the possible attribute value occurrences in it, declared as preferred and alternative labels, as shown in example below.

Example 1 (*)

ANSC EAD:

<name role="forma">rapsodia</name>
<name role="organico">pianoforte, orchestra</name>

Mapping :

<name> dc:description x

x rdf:Type skos:Concept
x skos:prefLabel V1
x skos: altLabel V2

In other cases it was necessary to actually create **new properties**. This was done in both fonds for the <geogname> element and its attribute @role (see AV table elements 23&25, and ETN table elements 34&35, with a "&" in column d). The geographic

location was specified by the attribute role, whose value specified whether the location in question was just a place, or a region or a country (in Italian Localita', Regione, Stato, respectively). Now this information is considered to be important for the "Where?" queries, and therefore new properties have been defined. The <geogname> element was mapped to ***dcterms:Spatial*** that may have ***ens:Place*** as a range. Then new properties that have ***ens:Place*** as a domain have been declared, and have been called ***ens:Place***, ***ens:Region*** and ***ens:State***. Finally, the mapping of the <geoname> element was done in a way that depends from the value of @role, as shown in example below:

Example 2 (&):

ANSC EAD

<geogname role="località">Catania</geogname>
<geogname role="regione">Sicilia</geogname>
<geogname role="stato">Italia</geogname>

Mapping

if	@role= località	then map <geoname> to	ens:Location
if	@role=regione	then map <geoname> to	ens:Region
if	@role=stato	then map <geoname> to	ens:State

Another case in which it was decided to define a **new property** was to distinguish the information contained in the attribute @authifilenumber (authority file number) from other identifiers that may occur in data and that are mapped to ***dc:Identifier***. For this purpose I have created a new property called ***ens:Authfilenumber*** that is subproperty of ***dc:Identifier***. (see AV table elements 51, with a "^" in column d), as shown in example below:

Example 3 (^):

ANSC EAD

@authfilenumber="00006231"

Mapping

ens:authfilenumber rdfs:SubpropertyOf dc:Identifier

Finally, as mentioned previously, for the hierarchical nodes that were **not translated** into instances of ***ore:Aggregations***, subclasses of ***ens:NonInformationResource*** called ***ens:ArchivalFond***, ***ens:ArchivalSeries*** and ***ens:ArchivalSubfonds*** were created.

It is important to note that for each of the new classes and properties that have been created, it should formally be stated to which namespace they belong to. The namespaces can also be defined by the institution. In the case of this thesis, I have used the Europeana namespace (abbreviated in ens:), and, for simplicity, I have omitted to declare completely the possible domains and ranges of each new property, and their relationships with the other classes.

4.2.5 Issue: Creating Instance of an Agent

This section describes the mapping of personal and corporate names that carry the attribute @role with further refinements such as performer, conductor, composer etc. and for which an instances of ***ens:Agent*** class were made.

If I want to say that Paganini was the "performer" in one concert, I create an Event that represents this musical performance, I then state that Paganini ***hasMet*** that Event, and I can state that the ***type*** of Paganini is "performer" in order to highlight his role. But this has the obvious problem that now Paganini is always of type "performer", even in the events in which he has been e.g. "conductor". To cover this case, the proper solution is to represent the Event as an Aggregation, and link to that Aggregation a Proxy for Paganini, and declare that the ***type*** of the Proxy is "performer". In this way Paganini is a "performer" **only in the context of that event**. In the context of a different event (such as the event where he was a conductor), then Paganini may have a different ***type***, i.e., "conductor".

In the ETN fond, I have applied the first version, where Agent always has ***dc:Type*** "performer", because of the nature of the material in this collection. In the AV fond, for each Event an ***ore:Aggregation*** was created. Furthermore, for each <persname> and <corpname> also a Proxy was created, linked to the appropriate Event Aggregation, and carrying the property ***dc:Type*** with the appropriate value. To the same Proxy is also linked ***dc:Description,*** which maps the <emph> element, carrying the information of the instrument that the person was playing).

4.2.6 Issue: Events Creation

In this section further clarification about the creation of events is provided. Only in the AV fond events were creates, since they exist in the original data. Events in this fond can be of different nature, for example concert season, musical performance, interviews, musical lessons etc. They are directly indicated by the element <event>, carry the date when it happened (in the element <date>) and carry the source and id of the event (<num id= "">). On the other hand, in EDM the Event class is a domain of property **ens:happenedAt** (the range of this property is ***ens:Place***) and also a domain of property **ens:occurredAt** (the range of this property is ***ens:TimeSpan***). For this reason for the <date> element related to the event in question (see table AV, element 96.), an instance of ***ens:TimeSpan*** is created and connected to the Event by ***ens:occuredAt***. Similarly, the element <geogname> (see table AV elements 23. and 59.) indicates where the event happened. For it, an instance of ens:Place is created and connected to ***ens:Event*** by ***ens:happenedAt*** .

This event is going to be further connected to other resources that participated in the same event. Therefore instances of classes ***ens:Agent*** (<persname>, <corpname>), ***ens:InformationResource*** and ***ens:PhysicalThing*** are to be connected with the event by the property **ens:WasPresentAt**.

4.2.7 Issue: Replacing Literals with Linked Open Data URIs

For the reason of representing geographical locations in ETN fond, I have decided to replace the literals representing geographical locations with the URIs of the geographical locations that those literals represent, and declared that those URIs are instances of the class ***ens:Place***. The URIs used in this case come from the *GeoNames* Ontology, which has over 6.2 million toponyms and make their unique URI available through a web service (Geonames, n.d.) This database is also a part of the Linked Open Data cloud, and the consequences for providing links to Linked Open Data together with the opportunities this step may bring, will be discussed in the Chapter 5 of this thesis. In this way, wherever it was found a location name in the EAD data, e.g. *Torino*, in EDM it was replaced by the appropriate URI representing this location in *GeoName*s (in this case http://www.geonames.org/3165524/torino.html).

Please note that this process happens only with the <geogname> element found in the ETN mapping table (see row 34).

4.3 Summary of the Validation Process and EDM

Before commenting on the Validation process and EDM I would first like to stress some of the limitations that may have influenced my work. As already stated, the analysis part was made with a developer's support on extracted DTDs from which elements were retyped in table, in order to reconstruct the path in which they are found in XML files. This method may result in errors or omissions done during analysis and retyping. Knowing XPath language would maybe provide with quicker and more accurate results.

Furthermore, because of the lack of through knowledge in EAD and archival description practice, some of my mapping decisions may seem inadequate to the archivist practitioners. Also in order to exploit maximum of the information ANSC EAD is carrying, it would be very useful to know searching behavior of the musician public, however because of the time limit in which this thesis has to be conducted, no such user study was performed. I have tried to conclude which information was perceived as more important by looking at the information offered in on-line presentations of these fonds.

Additionally, because of the lack of experience in dealing with RDF model and conceptual data modeling per sè, some of the solutions I have made may not be the best possible ones. Finally, Another limitation is the way of presenting. The only way I could express my mapping decisions is through plain text, which may have resulted in its readability. However, this was due to yet another lack in background knowledge of a formal language, through which my mapping decisions would possibly be expressed more clearly.

After performing this remodeling of data I would like to express few remarks on the process and the EDM model itself. The effort required in order to map ANSC EAD schema to EDM was not negligible. In addition to my lack of experience in conceptual modeling, there is currently a lack of examples for mapping to EDM schema and also the documentation on the EDM model itself is very scarce, which can be justified only by the fact that the model is still "young" and in a prototyping phase. To help a better understanding of the process, in the Figure 19 I am demonstrating the EDM graph corresponding to the EAD "branch" of the ETN fond, shown in Appendix 4.

What can be concluded from the work performed was that the model and its classes and properties are general, and sometimes too general for the purpose of presenting metadata of a particular institution. However ontology and RDF mechanism used in EDM allow retaining precision if considered as necessary, and some examples of how I achieved that were demonstrates in previous chapters. Therefore, it is possible to translate from EAD to the EDM model with minimal data loss. Still, in my opinion the EDM model has room for improvements, especially for the purpose of conforming to the well established practices that are used when presenting metadata of memory institutions.

For example, a reference element that provides a citation and/or electronic link for a **published work** is usually encoded in EAD by the <bibref> element. This information was represented in EDM by the ***ens:Realizes*** property, whose domain are instances of ***ens:InformationResource*** holding the metadata on the particular publication described (see AV table, elements 26-43). In this specific case the attribute @role specified that the

value of <bibref> was referring to **published material**, and therefore this information was saved by mapping @role to ***dc:Description*** that was attached to instance of ***ens:InformationResource*** in question. If no such description was available, the information of the original tag (<bibref>) would have been lost. For this and other reasons one may identify when mapping archival, library or museum domain metadata to EDM, application profiles should be developed for each of those domains and included in the EDM.

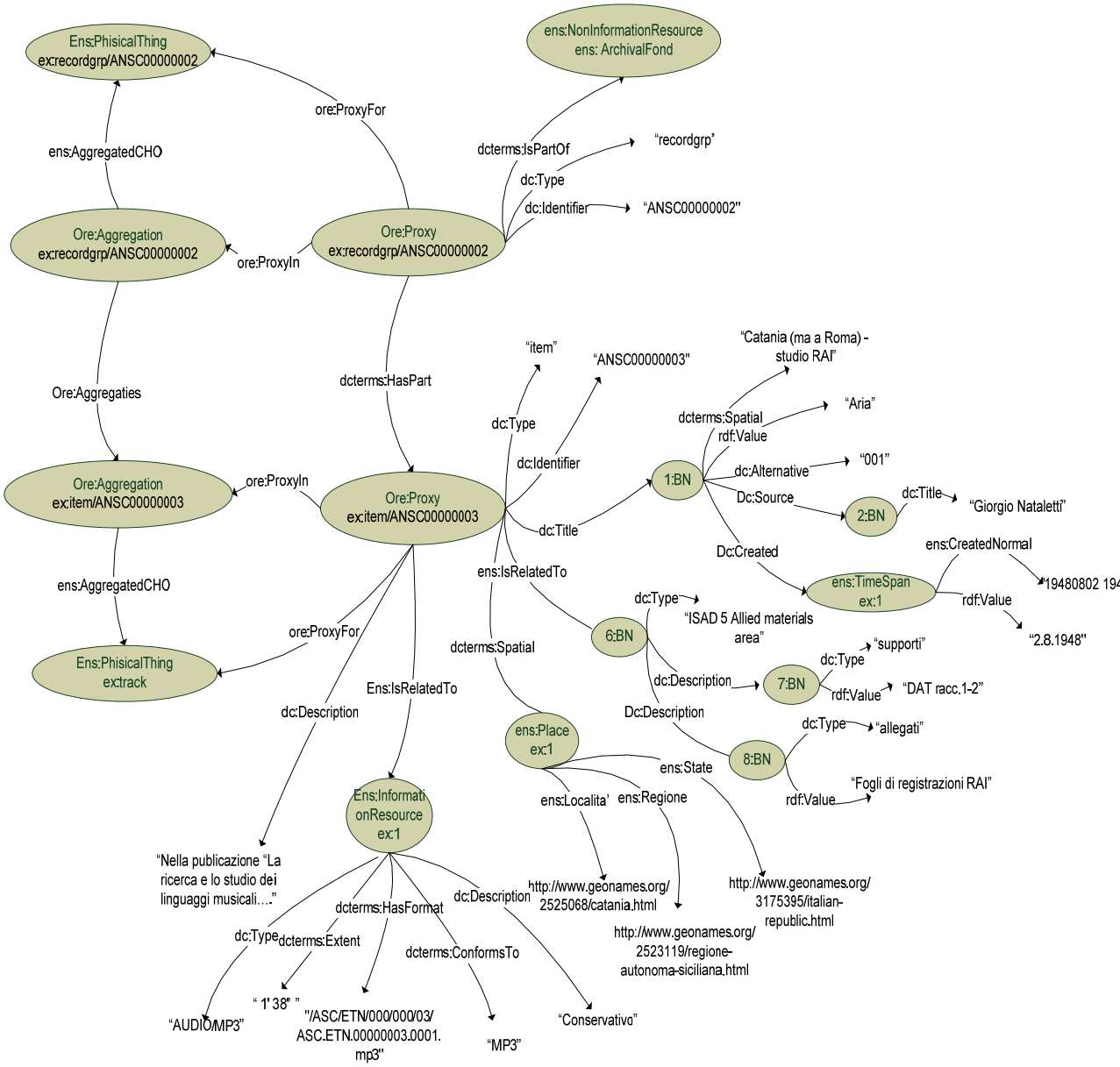

Figure 19: Example of the EDM modeled "branch" from Ethnomusicology fond
(only item level developed for the readability; ens:ArchivalFond node developed in figure 18)

4.4 Discussion on the Validation Results

In this section I will analyze the work I have conducted and previously described on mapping the two fonds provided by the Accademia di Santa Cecilia (described by EAD data) to the EDM model. This analysis and discussion will be done having in mind the questions and problem areas raised in the theoretical framework, i.e. what does general user population find problematic with current online finding aids, and what are their preferences for access: In the literature review four main categories of issues were outlined:

1) **archival terminology used in on-line presentation,**

2). **structure in which the presentation is organized and expressed (hierarchical structure),**

3) **access possibilities that support mainly browse and not search function,**

4) **unawareness of users of particular archival holdings, caused by the EAD structure in which a description can be found, which hampers Web crawlers from reaching information and union archival searches.**

4.4.1 Structure Issues

From the comments made throughout this thesis, it should be clear that archival description is usually based on arrangement of materials according to provenance or original order of records and commonly instantiated by means of a unique EAD XML file which combines the hierarchical structure elements with the content elements, without a clear distinction between the two. This fact makes it difficult to determine how to access and exchange a specific subset of data without navigating the whole hierarchy or without losing meaningful hierarchical relationships. (Silvello, 2011). This in turn influences the novice users experience when accessing archive on-line and searching for the wanted material, as they usually want a quicker way to reach the desired information and very

often do not have the skill (or the patience) for following the usual top-down approach through the hierarchy. (Shier, 2004)

The structural mapping shown in section 4.2.1 and Figure 17 demonstrated that when converting to EDM the nodes that were arranged in the hierarchy of the EAD data were extracted and converted to separate objects of their own. In some sense, the hierarchy has been flattened and all the objects become equally accessible, regardless of whether they were on top or at the bottom of the original hierarchical data. It was also shown that by relating the Proxies of those extracted objects with the ***dcterm:HasPart*** and ***dcterms:isPartof*** properties, the original hierarchical relations and the context in which the information can be found were retained

This change allows a user who makes a query, to "land" directly at any of the conceptual levels and from that point to conduct a top-down or bottom–up navigation, or actually any kind of movement through described material, depending on the need.

As noted in section 4.2.1 not all the nodes that are in the original hierarchy are extracted as separate object (i.e. no Aggregation is created for them). The reason being that they do not carry enough descriptive information but carry only the concept that the lower level "are together". For example, level "fond" of ethnomusicology can be understood as the "room" in the archive where all other "recordgroup" levels are grouped, as the only metadata it carries is that of its physical location and the title. Therefore, I have decided to attach to each of the lower levels nodes (in this case "recordgroup") the information that they belong to the same fond. This means that the notion of belonging to the same source is kept, without actually having to go through that source to reach the wanted one. This may as well influence the final display but we have not investigated how the mapping decisions can affect the presentation of information. Still, this decision should result in reducing the levels through which the user accesses information if following the top-down approach.

Finally, practically all users make use of search engines to search for relevant material in the Web, and they usually use access points (e.g. personal, corporate, and geographical

names, topical subjects etc.) that very often in archival descriptions are encoded at the lower levels of the hierarchy, meaning they are buried too deep within the XML document to trigger the Web page searches (Hill, 2004; Kiesling, 2001). By mapping to EDM, extracting the levels and making them as equals should also make the information of the lower levels more accessible to the Web crawlers, therefore increasing the chance for the archival material to be retrieved by Web search. This should also directly influence the increase in visibility of archival holdings.

4.4.2 Terminology Issues

As noted by Gilliland-Swetland (2001) users find confusing the administrative information that is woven throughout the finding aids. By mapping to EDM I have taken out all the information that may not be relevant to the user when discovering resources and reading their description, such as administrative metadata encoded in <processing> element. This should result in clearer presentation of information.

Edison (2002) points out that EAD was not built with users in mind, but was built by the archivists for the archivists, which can be noted in the encoding terminology used in EAD elements. This, in turn influenced also the final presentation, where the use of archival jargon was one of the main problems for the novice users, as seen in the section 2.2.1 of Chapter 2. Using the archival jargon (e.g. Scope and Content, Container List, Extent, Arrangement, Bulk, and so on) are essential categories or descriptors for archival collections on the input side. However, when used for output, which is the display, made in turn problem to the novice ones who had to firstly interpret this jargon in order to find wanted information. The advantage obtained with this mapping is that now the effort to interpret and translate archival terms and concepts has been moved from the many users to the one person who actually conducts the mapping and the "translation" to a more widely understood jargon, i.e. the one of EDM. As shown in Appendix 3, EDM is a more general language and was created with the goal of presenting cross-domain descriptions. The core of EDM is based upon Dublin Core elements that have a commonly understood semantics, but also the other classes and properties carry the semantic that is domain

independent, and therefore should be more intuitive for the users than the one originating in archival practice. This can help eliminate the problems that general users have when dealing with archival terminology.

Furthermore, as noted by Kim (2004) the terminology used in different institutional on-line finding aids is inconsistent, both in wording and in meaning. This can cause a problem to the user who reaches several different archival repositories. An inherent problem with every language, whether archival or not, is that it can be interpreted in different ways. In order to prevent this language-based confusion, clear and precise language should be used and, and it would be desirable that the same language could be used across institutions to make them more accessible to users. (Johnston, 2008) Now, it is still not yet clear which exact terminology will be used as output of EDM, but what is certain is that if all different uses of EAD were mapped in the same way to EDM, it would result in a uniform way of representing information across repositories, which should result in a reduction of ambiguity and confusion within members of the general public.

4.4.3 Union Search

Further to the last point made in the previous section, it can be added that EDM as a model is allowing interoperability at the repository level. (Chan, Zeng, 2006) The model allows implementing Web server and performing cross-collection searching of all the EAD encoded descriptions mapped to it. Furthermore, it allows not only cross archival but also cross domain search, as demonstrated in Appendix 3.

4.4.4 Search Issues

The most useful access points for accessing the archival records are: subject , personal name, geographical location, genre and event (as shown in section 2.2.3.1). EDM directly supports search through all of these parameters, and how I dealt with some of them was shown in **example 2** and section 4.2.5 of this chapter. What is the added value regarding

the search possibilities is the conceptual modeling that EDM uses to represent the data i.e. the use of the SKOS naming features. For example, in the AV fond some of the objects were described with having the author “Albéniz, Isaac”, while other with having the author “Albeniz,Isaac”. Now, it is pretty clear that the author in question is the same, with the difference in accent. But this makes impossible to retrieve all the work from this author using just one of those two names. If, on the other hand I use the SKOS concept, I can replace all those literals by a URI U and declare that U has “Albéniz, Isaac” as a preferred label and “Albeniz,Isaac” as an alternative label. Therefore, when the user searches using any of the two labels, the system is able to determine that he/she wants the work of U, and will returns all the work by this author. Furthermore, since the SKOS relationships are expressed in RDF, this brings into the picture the inference mechanism of RDF, which can result in the discovery of new relations that were not encoded in original data. For example, if the data is presented in an event-centric way, there is a new relation between two Agents (e.g. performers) that Were Present at the same Event.

4.4.5. Opening the Borders

Another added value that the EDM model would bring to the general user accessibility, is the possibility of adding user generated description by means of the property ***ens:UserTag***. User tagging, also known as collaborative tagging (Golder & Huberman, 2006; Macgregor & McCulloch, 2006), social tagging (Tennis, 2006), or social bookmarking (Hammond, Timo, Ben & Joanna, 2005) is tagging done by the “users” of search services i.e., by those whose participation in the resource discovery or information retrieval process has historically been limited to the expression of information needs and construction of search queries. (Furner, 2007)

This point was not included in the theoretical framework as a crucial one, but it has been certainly discussed in archival circles as a desirable functionality. Allowing users to augment the online finding aids with their own tags, archivists could exploit their roles as mediators and producers of knowledge, creating a powerful tool for description, revision, reference, and research. (Light & Hyry, 2002)

"For enhancing description in a finding aid, annotations could capture increasing amounts of detail about a collection or offer different perspectives on it. For instance, archivists and researchers might call attention to specific items within folders. They might elaborate on what can be found in a series, section, or folder. This would promote discovery by augmenting the existing form of access or by offering alternative descriptive language that may lead researchers to places they might have overlooked otherwise." (Light & Hyry, 2002, p.228) Furthermore, as users with similar interests tend to have a shared vocabulary, tags created by one user are useful to others, especially those with similar interests as the tagger's. (Wu, Zubair & Maly, 2006). Every archivist, while appraising, arranging and describing material gives a particular *view* to it, and this *view* would not be disrupted by the addition of users' tags, as EDM utilizes **Proxies** to separate the *view* of the archivist from those generated by users on the same material.

4.5. Chapter Summary

In this chapter I have demonstrated the work on validation of the general method for mapping EAD to EDM, by showing how it was applied to Ethnomusicology and Audio Video Fonds of Accademia Nazionale di Santa Cecilia. I have explained in which way the mapping was conducted, and demonstrated some of the decisions made through the examples. Finally, I have included a discussion on how this mapping can improve the accessibility of archival data for the general user public. Conclusions to the research questions, including summary of the most significant points provided in this chapter, and directions for future research will be presented in the next chapter.

CHAPTER 5: Conclusion

This chapter addresses directly the research questions by summarizing the points made in section 4.5 of the previous chapter. Furthermore, I will include some additional remarks that build up to the problem in question. Finally, I will provide some recommendations for future research directions.

5.1. Conclusion to the Research Question

The research question :

- **Would transforming EAD encoded archival descriptions in EDM bring improvements to on-line access for general archival public?**

As the results of the research project conducted, I have concluded that transforming EAD encoded archival description into the EDM model would bring certain improvements for what concerns making: single archives more accessible to general users and making multiple archives more accessible to non-experts via this model. These improvements can be summarized as follows:

- Archival terminology is one of the main problems to novice users, but after mapping it was translated to the more general language used in EDM, therefore these users would not have to experience such difficulties anymore. Furthermore, inconsistency of the language used in different institutional web sites, that made archival research confusing, would also be solved if all those different institutions would map their EAD data to EDM in a uniform way. Novice user would use a single, simpler language, instead trying to learn many different institutional ones.

- General user that found the hierarchical structure of finding aids confusing, after the mapping to EDM would not have to deal with it anymore. What was usually

available for discovery through a top-down approach, using EDM as a query language can now be reached directly and from there a user can go in any direction possible.

- General user that accesses the archival holdings through Web search engines has better chance to find the wanted material if its description is encoded in EDM. In EAD the relevant information is usually buried too deep in the file for Web crawlers to index them, as they usually read only the header information. In EDM, the information from the lower levels is extracted and as equally accessible to search engines as the one from the upper ones. Furthermore, the EDM language allows the possibility of conducting union archival search, if the separate archives map to EDM in a uniform way.

- Search possibilities may also be improved now that the data is conceptually modeled, by exploiting the links between objects and their contextual resources, which can be described in a richer way (e.g. labeling of concepts, semantic relations between concepts).

- Finally, it allows the enrichment of the description from the side of the users who can contribute to it by user tagging features, which could result in better accessibility and in an enrichment of the archival material.

The conclusions to the research question are the result of the research design applied in this study and are made on a theoretical basis. In order to prove them some further research should be conducted, which is discussed in section 5.3.

5.2. Further Remarks

Previously, I have tried to answer to the research question by making several arguments to show which improvements would be achieved at the ***model level***. However, one of the important conclusions is that EDM allows some improvements at the ***data value level*** as

well. In my opinion, the *possibility* and *potential* that this model allows on the data value level may be even bigger than those made by just remodeling the data in EDM.

One important possibility is brought by the replacement of literals with URIs from the Linked Open Data cloud, and enriching existing data in this way. One example of such possibility has been shown in ETN fond (section 4.3.6) where ***GeoNames*** URIs were used to replace literals of the geographic names. This step would result in answering directly to the specific request made by one of the most important groups of archival users which are genealogists. For the genealogists, finding information through geographic location names is the second most important access point, after personal names. "Archives usually organize and index records by the name a locality had when the records were created. Therefore genealogists must be able to link current place names with former place names, and vice versa." (Duff & Johnson, 2003, p. 86). Using *GeoNames* URIs would allow this linking (e.g. user may search for the literal Saint Petersburg and still retrieve documents that were indexed by Leningrad) but also it allows multilingual search (e.g. it doesn't matter if user will search Saint Petersburg or Санкт-Петербург, retrieval would be the same). Furthermore, genealogists often know the general area where their people lived, but not the exact location or the name of the town, so they need to consult reference tools such as maps to obtain this information (Duff & Johnson, 2003). *GeoNames* resources describe geographical locations with providing the Google map of the specific one, and therefore allowing genealogists to find what they want with as less effort as possible.

Using *GeoNames* is just one example I have used in order to demonstrate how reusing already available Web resources would answer some of the user needs. Still, there are many more reference value vocabularies that can be used, for example: The Virtual International Authority File (VIAF), Getty Union List of Artist Name (ULAN), The Getty thesaurus of Geographic Names (TGN), Dewey Decimal Classification in Linked Data (DDC), Universal Decimal Clasification (UDC), Classification System for Art and Iconography (Iconclass), DBpedia knowledge base and so on. (Isaac, n.d.)

Furthermore, as it was shown in the Introduction of the thesis, archival description is usually based on its fundamental principle which is the "principle of provenance" that dictates that records of different origins (provenance) should be kept separate to preserve their context. This means that archival researcher needs to access several fonds in order to find material of interest, that may be connected in some way, but are kept and described as parts of separate fonds. The Linked Open Data is a way of vastly expanding the benefits of search, by helping users discover contextually related materials. Creating links between archival collections and other sources is crucial as archives relating to the same people, organizations, places and subjects are often widely dispersed, and using LOD URIs would bring these together intellectually and new and surprising discoveries could be made about the life and work of an individual or the circumstances surrounding important historical events. (Use Case LOCAH, n.d.)

5.3. Implications for Further Research

The research conducted for the purpose of this thesis has resulted in discussion on the improvements that transforming EAD encoded archival description into EDM will bring to the general user population. This discussion is, however on the theoretical level.

This study could be taken as a starting point for more complex and ambitious research that could examine these findings in real life cases, by compare the search success when querying the same set of archival data, described both in EAD and EDM, in order to confirm or refute my findings.

In addition, a retrieval study could be conducted on popular Web search engines, in order to investigate whether there is an improvement in their indexing of archival descriptions when encoded in EDM as compared to the same descriptions when encoded in EAD.

Moreover, an ambitious research could be conducted in order to develop a single mapping solution to EDM that would cover all the idiosyncratic uses of EAD in order to provide a high quality union archival search.

As seen in Chapter 2, researches are still very scarce when it comes to discovering the newly emerging, on-line public and their preferences for the discovery of archival resources. A bigger scale research should be conducted in order to obtain more information on this issue, preferably dealing with users log analysis. Also, researches on on-line access preferences of non-expert members of different groups of archival user public (e.g. genealogists, historians, academic users, professional users etc) should be further examined, as the existing ones that I have found for the purpose of this study, mainly dealt with physical archival access.

The results of such studies should be applied for the purpose of building a product for EDM encoded archival data, which is a user interface that could provide intuitive access possibilities. Studies should be conducted on different groups of users to evaluate the effectiveness of the new interface.

Finally, the utilization of Linked Open Data URIs should be examined more closely, in order to understand the full capabilities of Linked Data and RDF inference.

Altman, B. & Nemmers, J. R. (2001) The Usability of On-line Archival Resources: The POLARIS Project Finding Aid. *American Archivist*, 64. 121-131.

Antoniou, G.&van Harmelen, F. A (2008) *Semantic Web Primer,* 2nd Edition (Cooperative Information Systems), The MIT Press.

Baca, M (2003) Practical issues in applying metadata schemas and controlled vocabularies to cultural heritage information. *Cataloging & classification quarterly*. 36(3). 47–55.

Bantin, P.C. (2001). Strategies for Managing Electronic Records: Lessons Learned from the Indiana University Electronic Records Project. retrieved 20.06.2011 http://www.indiana.edu/~libarch/ER/bantin-saa2001.pdf

Bates, M. (1989) The design of browsing and berrypicking techniques for the on-line search interface. *Online Review* 13(5). 407-431

Bates, M. J., Wilde, D. N. & Siegfried, S. (1993) An Analysis of Search Terminology Used by Humanities Scholars: The Getty Online Searching Project report number 1. *Library Quarterly*, 63(1) 1-39.

Berners-Lee, T. (1999) *Weaving the Web*. London: Orion Business Books

Bibliomediateca, who are we (n.d) retrieved 26.07.2011 from http://bibliomediateca.santacecilia.it/bibliomediateca/cms.view?munu_str=0_5_0&numDoc=331

Carpenter, B. & Park, J. (2009) Encoded Archival Description (EAD) Metadata Scheme: An Analysis of Use of the EAD – Headers. *Journal of Library Metadata*, 9(1). 134.

Chan, L. M & Zeng, M. L. (2006) Metadata Interoperability and Standardization - A Study of Methodology, Part I. *D-Lib Magazine*, 12(6)

Coats, L. R. (2004) Users of EAD - Finding Aids: Who Are They and Are They Satisfied? *Journal of Archival Organization*, 2(3).25.

Combs, M., Matienzo, M.A., Proffitt, M.& Spiro, L. (2010) Over, Under, Around, and Through: Getting Around Barriers to EAD Implementation. OCLC Online Computer Library Center, Inc. Retrieved 26. 06. 2011 from www.oclc.org/research/publications/library/2010/2010-04.pdf

Concordia, C., Gradmann, S. & Siebinga, S. (2009) Not (just) a Repository, nor (just) a Digital Library, nor (just) a Portal: A Portrait of European as an API. International *Federation of Library Associations and Institutions.* 36(1). pp. 61–69. (http://dx.doi.org/10.1177/0340035209360764)

Conway, P. (1986) Facts and Frameworks: An Approach to Studying the Users

Conway, P. and Partners in Research (1994) *Improving Access to the Nation's Archives*. Pittsburgh. Archives and Museum Informatics.

Cox, R. J. (1998) Access in the Digital Information Age and the Archival Mission: The United States. *Journal of the Society of Archivists*, 19(1). 30-31.

Daniels, M.G. & Yakel, E (2010) Seek and You May Find: Successful Search in Online Finding Aid Systems. *American Archivist* 73(2).535–568.

Dearstyne, B. (1987) What is Use of Archives? A Challenge for the Profession. *American Archivist*. 50

Definition of the Europeana Data Model elements Version 5.2.1, (2011) Europeana v1.0. retrieved 20.06.2011 from http://www.version1.europeana.eu/c/document_library/get_file?uuid=aff89c92-b6ff-4373-a279-fc47b9af3af2&groupId=10605

Development of the Encoded Archival Description DTD, (n.d.) retrieved 26.07.2011. from http://www.loc.gov/ead/eaddev.html

Doerr, M. & LeBoeuf, P. (2007): Modelling Intellectual Processes: the FRBR - CRM Harmonization. In: C. Thanos, F. Borri, and L. Candela (eds.): Digital Libraries: R&D, LNCS 4877, pp. 114–123, 2007. (First DELOS Conference on Digital Libraries, February 2007 Tirrenia, Pisa, Italy.)

Doerr, M., Gradmann, S., Hennicke, S., Isaac, A., Meghini, C. & van de Sompel, H. (2010) The Europeana Data Model (EDM) IFLA 149. Information Technology, Cataloguing, Classification and Indexing with Knowledge Management. retrieved 20.06.2011 from http://www.ifla.org/files/hq/papers/ifla76/149-doerr-en.pdf

Duff, W. M. & Stoyanova, P. (1998) Transforming the Crazy Quilt: Archival Displays from a User's Point of View. *Archivaria* 1(45)

Duff, W. M. & Harris, V.(2002) Stories and names: Archival description as narrating records and constructing meanings. *Archival Science*, 2 (3-4), 263-285.

Duff, W.M. & Johnson, C. A. (2003) Where Is the List with All the Names? Information-Seeking Behavior of Genealogists. *The American Archivist* 66(1). 79-95.

Dushay, N. & Hillmann, D. (2003) Analyzing Metadata for Effective Use and Re-use 2003 Dublin Core Conference, Seattle, WA.

EDM Archives meeting – Mapping Archival Data to the EDM, minutes (2010) Europeana Documentation, retrieved 20.07.2011 at http://europeanalabs.eu/wiki/WP1CommunityMeetingArchives

Eidson, M. Y. (2002) Describing Anything That Walks: The Problem Behind the Problem of EAD. *Journal of Archival Organization*, 1(4)

Encoded Archival Description application guidelines : version 1.0, (1999) Chicago, IL: Society of American Archivists.

Encoded Archival Description Tag Library , Version 2002 retrieved 26.06.2011 from http://www.loc.gov/ead/tglib/

Europeana Data Model Primer. Europeana v1.0. (2010) retrieved 20.06.2011 from http://www.version1.europeana.eu/c/document_library/get_file?uuid=718a3828-6468-4e94-a9e7-7945c55eec65&groupId=10605

Europeana, Background (n.d.), retrieved 20.06.2011 from http://www.europeana.eu/portal/aboutus_background.html

Europeana, Think Culture (n.d.), retrieved 20.06.2011 from http://www.europeana.eu/portal/aboutus.htmll

Europeana, v.1.0 Project (n.d.), retrieved 20.06.2011 from http://version1.europeana.eu/web/europeana-project

Fox, M. J. (2001) Stargazing: Locating EAD - in the Descriptive Firmament," *Journal of Internet Cataloging*, 4(3) (2001). 61.

Furner, J. (2007) User tagging of library resources: Toward a framework for system evaluation, 73RD IFLA General Conference and Councile,

GeoNames Ontology retrieved 26.07.2011 from http://www.geonames.org/ontology/documentation.html

Gilliland-Swetland, A. J. (1998) An Exploration of K-12 User Needs for Digital Primary Source Materials. *The American Archivist*, 61(1).136-157.

Gilliland-Swetland, A. J. (2000) Introduction to Metadata: Setting the Stage. In Murtha Baca, ed. Introduction to Metadata: Pathways to Digital Information. Los Angeles, Calif.: Getty Research Institute.

Gilliland-Swetland, A. J. (2001) Popularizing the Finding Aid. *Journal of Internet Cataloging,* 4(3/4) 199-225.

Gilliland-Swetland, A. J. (2002) Enduring paradigm, new opportunities. Washington DC: Council on Library and Information Resources.

Golder, S. A.&. Huberman B. A. (2006) Usage patterns of collaborative tagging systems. Journal of Information Science, 32(2). 198–208.

Gruber, T. R. 1993. A translation approach to portable ontology specifications. Knowledge Acquisition, 5: 199-220

Hammond, T., Timo, H., Ben, L., Joanna, S. (2005) Social Bookmarking Tools (I). *D-Lib Magazine* 11(4)

Haworth, K.M. (2001) Archival Description: Content and Context in Search of Structure. *Journal of Internet Cataloging*, 4(3). 7.

Heery, R. & Patel, M. (2000) Application profiles: mixing and matching metadata schemas. Ariadne 25.

Hennicke, S. (2010) Mapping Archival Data to the EDM, Community Meeting: Archives, Berlin, 26 April, Presentation Slides provided via email by the author.

Hill, A. (2004) Serving the invisible researcher: Meeting the needs of online users. *Journal of the Society of Archivists*, 25(2).139-148.

Hillmann, D. Dushay, N. & Phipps,J. (2004) Improving metadata quality: augmentation and recombination. in Proceedings of the 2004 international conference on Dublin Core and metadata applications: metadata across languages and cultures (Shanghai, China: Dublin Core Metadata Initiative, 2004), 7:1–7:8.

Hostetter, C. J. (2004) Online Finding Aids: Are They Practical? *Journal of Archival Organization*, 2 (1/2). 117–145.

Hoyer, T.P., Miller, S. & Pollock A (2001) Consortial Approaches to the Implementation of Encoded Archival Description (EAD): The American Heritage Virtual Archive Project and the Online Archive of California (OAC). *Journal of Internet Cataloging*, 4(3/4) (2001): 113-136. Reprinted in Encoded Archival Description on the Internet, Daniel V. Pitti and Wendy M. Duff, eds. (New York: Haworth Information Press, 2001) International Council on Archives, ISAD(G): General International Standard Archival Description, Ottawa, 1994

Isaac, A. (n.d) Use case Europeana; retrieved 26.07.2011 from http://www.w3.org/2005/Incubator/lld/wiki/Use_Case_Europeana

Isaak, (n.d.) On practical aspects of enhancing semantic interoperability using SKOS and KOS alignment, Presentation Slides ISKO UK Meeting, July 21, London, retrieved 26.06.2011 from http://www.iskouk.org/presentations/isaac_21072008.pdf

ISAD(G):General International Standard Archival Description Second Edition ,Ottawa 2000

Johnston, R. D. (2008) A Qualitative Study of the Experiences of Novice Undergraduate Students with Online Finding Aids. (Master dissertation) Retrieved from the Dissertation and Thesis Database: http://dc.lib.unc.edu/cdm4/item_viewer.php?CISOROOT=/s_papers&CISOPTR=1122&CISOBOX=1&REC=15

Kiesling, K. (2001) Metadata, metadata, everywhere–but where is the hook?. OCLC Systems & Services 17(2). 84–88.

Kim, J. (2004) EAD Encoding and Display: A Content Analysis. *Journal of Archival Organization.* (2)3. 41-55.

Landis, W. C. (1995) Archival Outreach on the World Wide Web. *Archival Issues*. 20(2). 129-147

Landis, W.(2002) Nuts and Bolts: Implementing Descriptive Standards to Enable Virtual Collections. *Journal of Archival Organization*, 1(1). 82-83.

Library of Congress Encoded Archival Description Best Practices (2008) Library of Congres

Light, M & Hyry, T (2002) Colophons and Annotations: New Directions for the Finding Aid. *The American Archivist* 65(2).216-230.

Lytle, R. H. (1980) Intellectual Access to Archives: Provenance and Content Indexing Methods of Subject Retrieval. *The American Archivis*t,

Macgregor, G. & McCulloch, E. (2006) Collaborative tagging as a knowledge organisation and resource discovery tool. *Library Review* , (55)5. 291–300.

Maher, W. J. (1986) The Use of User Studies. Midwestern Archivist ,11(1) (1986): 15

March, S and Smith, G. (1995). Design and Natural Science Research on Information Technology. Decision Support Systems, 15, 251–266.

Meghini, C. (n.d) Web Architecture. Lecture Slides. Document posted in IVA on-line learning environment at http://iva.htk.tlu.ee

Meghini, C., Isaac, A., Gradmann, S., Schreiber, G. et al. (2010) The Europeana Data Model. ECDL Workshop on Very Large Digital Libraries Glasgow, September 10, 2010 Document posted in IVA on-line learning environment at http://iva.htk.tlu.ee

Menne-Haritz, A. (2001) Access—the reformulation of an archival paradigm. *Archival Science*, 1(1), 57–82.

Menne-Haritz, A. (2008) Archives on the Internet: sharing data across domains. Presentation held during the European Digital Library conference in Frankfurt on the 31st of January 2008. Retrieved 20.03.2011 from http://www.theeuropeanlibrary.org/portal/organisation/cooperation/archive/edl

Milestone Europeana Documentation, retrieved 26.07.2011 from Google in PDF

Minutes: Mapping Librarian Data to the EDM (2010), Europeana Office, Europeana Documentation retrieved 20.07.2011 from http://europeanalabs.eu/wiki/WP1CommunityMeetingLibraries

OAC Best Practices Guidelines for Encoded Archival Description (2005) OAC Working Group. retrieved 26.06.2011 from http://www.cdlib.org/services/dsc/contribute/docs/oacbpgead_v2-0.pdf
of Archives. American Archivist, 49.393-407

Olensky, M. (2010) Semantic interoperability in Europeana: An examination of CIDOC CRM in digital cultural heritage documentation; *TCDL Bulletin*, 6 (2)

Open Archives Initiative (n.d) retrieved 20.07.2011 from http://www.openarchives.org/

Pearce-Moses, R. (2005). A glossary of archival and records terminology. Chicago: Society of American Archivists

Pitti, D. V. (2001) Encoded Archival Description: An Introduction and Overview, ESARBICA Journal, 20. 71-80

Pitti, D. V., & Duff, W. M. (2001) Encoded Archival Description on the Internet, Binghamton, NY: Haworth Information Press. Also published as Journal of Internet Cataloging. 4 (3/4).

Prom, C. J. (2003) Reengineering archival access through the OAI protocols. *Library hi tech* 21(2).199–209.

Prom, C. J. (2004) User Interactions with Electronic Finding Aids in a Controlled Setting. *The American Archivist*, 67(2). 234-268.

Prom, C. J.& Habing, T.G. (2002) Using the open archives initiative protocols with EAD Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries. 171–180.

Pugh, M. J. (1982) The Illusion of Omniscience: Subject Access and the Reference Archivist. *The American Archivist*, 45(1).33-44.

RDF Primer-W3C Recommendation (2004) retrieved 26.06.2011 from: http://www.w3.org/TR/rdf-syntax/

Ribeiro F. (2001) Archival science and changes in the paradigm. Archival Science, 1(3). 295-310.

RLG Best Practices Guidelines for Encoded Archival Description. (2002) Research Libraries Group , RLG EAD Advisory Group

Rosenbusch, A. (2001) Are Our Users Being Served? A Report on Online Archival Databases. *Archives and Manuscripts*. 29. 44-61.

Ruth, J. E. (2001) The Development and Structure of the Encoded Archival Description (EAD) Document Type Definition. *Journal of Internet Cataloging* 4(3). 27.

Schier, W. (2006) First Entry: Report on a Qualitative Exploratory Study of Novice User Experience with Online Finding Aids. *Journal of Archival Organization* 3(4). 49-85.

Shaw, E. J. (2001) Rethinking EAD: balancing flexibility and interoperability: Interoperability. *New review of information networking*, 7. 117–131.

Silvello, G. (2011) A Set-Based Approach to Deal with Hierarchical Structures. (Doctoral Disertation). Available via email by Carlo Megini

Standard MAG - Versione 1.5, (2004) retrieved 26.07.2011 from http://www.iccu.sbn.it/opencms/opencms/en/main/standard/metadati/pagina_103.html

Statement of Principles Regarding Archival Description (1992) International Council on Archives Ad Hoc Commission on Descriptive Standards, Madrid.

Styles, R., Ayers, D & Shabir, N (2008) Semantic MARC, MARC21 and the Semantic Web. WWW 2008 17th International World Wide Web Conference.

Tarrant, D., O'Steen, D., Brody, T., Hitchcock, S., Jefferies, N. & Carr, L. (2009) Using OAI-ORE to Transform Digital Repositories into Interoperable Storage and Services Applications, *Code4Lib Journal*, 6

Tennis, J. T. (2006). Social tagging and the next steps for indexing. Advances in classification research, Vol. 17: Proceedings of the 17th ASIS&T SIG/CR Classification Research Workshop (Austin, TX, November 4, 2006)

The State of State Records: A Status Report on State Archives and Records Management Programs in the United States (2007) Council on State Records. retrieved 20.06.2011 from http://www.statearchivists.org/reports/2007-ARMreport/StateARMs-2006rpt-final.pdf

Theodoridou, M. & Doerr, M.(2001) Mapping the Encoded Archival Description DTD Element Set to The CIDOC-CRM, Technical Report 289, ICS-FORTH retrieved 26.07.2011 from http://www.regnet.org/members/demo/covax/files/2001.TR289_mapping_of_the_encoded.pdf

Thurman, A. C. (2005) Metadata standards for archival control: an introduction to EAD and EAC. *Cataloging & classification quarterly*, 40(3). 183–212.

Tibbo, H. R. (2002) Primarily history: historians and the search for primary source materials. in Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries, JCDL '02 (Portland, Oregon, USA: ACM, 2002), 1–10.

Use Case LOCAH, retrieved 26.07.2011 from http://www.w3.org/2005/Incubator/lld/wiki/Use_Case_LOCAH

Vaishnavi, V. K. & Kuechler Jr.,V. (2007) Design Science Research Methods and Patterns: Innovating Information and Communication Technology, 1st ed. Auerbach Publications.

Wu, H., Zubair, M. & Maly, K. (2006) Harvesting social knowledge from folksonomies. Proceedings of the seventeenth conference on Hypertext and hypermedia, 111–114.

Yakel, E. & Torres, D. A (2003) AI: archival intelligence and user expertise. *American Archivist*, 66(1). 51–78.

Yakel, E. (2002) Listening to users. *Archival Issues,* 26(2). 11–127.

Yakel, E. (2003) Archival representation. *Archival Science*, 3(1).1-25.

Yakel, E. (2004) Encoded Archival Description: Are Finding Aids Boundary Spanners or Barriers for Users?. *Journal of Archival Organization* 2(1). 63-77.

Zhou, X. (2007) Examining Search Functions of EAD Finding Aids Web Sites. *Journal of Archival Organization,* 4(3). 99-118.

Appendices:

Appendix 1

Sample Encoded EAD Record (taken from: Thurman, 2005)

<?xmlversion="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE ead PUBLIC "+//ISBN 1-931666-00-8//DTD ead.dtd (Encoded Archival Description (EAD) Version 2002)//EN" "../shared/ead/ead.dtd>
<ead>
<eadheader audience="internal" countryencoding="iso3166-1" dateencoding="iso8601" langencoding="iso639-2b" repositoryencoding="iso15511">
<eadid countrycode="us" mainagencycode="xx-x" publicid="-//us::xx-x//TEXT us::xx-x::f24.sgm//EN">Basham Kelly papers</eadid>
<filedesc>
<titlestmt>
<titleproper>Guide to the Basham Kelly papers, 1936-1988</titleproper>
<author>Collection processed by Judith Morgan, finding aid prepared by Diana Elizabeth</author>
</titlestmt>
<publicationstmt>
<publisher>University Archives, Rodgers Library, Bluegrass State University.</publisher>
<date>© 1992</date>
</publicationstmt>
</filedesc>
<profiledesc>
<creation>Finding aid encoded by Richard Cooper,
<date>2004.</date></creation>
<langusage>Finding aid is written in
<language>English</language></langusage>

<descrules>APPM used for description; AACR2r used for descriptive headings;
LCSH used for subject headings.</descrules>
</profiledesc>
</eadheader>
<frontmatter>
<titlepage>
<titleproper>Guide to the Basham Kelly Papers.</titleproper>
<num>MS-F24</num>
<publisher>University Archives<lb>Rodgers Library<lb>Bluegrass State University
<lb>Danville, Kentucky</publisher>
<list type="deflist">
<defitem>
<label> Processed by:</label>
<item>Judith Morgan</item>
</defitem>
<defitem>
<label> Finding aid prepared by:</label>
<item>Diana Elizabeth</item>
</defitem>
<defitem>
<label> Encoded by:</label>
<item>Richard Cooper</item>
</defitem>
</list>
&copy 1992 Bluegrass University. All rights reserved.
</titlepage>
</frontmatter>
<archdesc level="collection" relatedencoding="MARC">
<did>
<head>Descriptive Summary</head>
<origination label="Creator"><persname encodinganalog="100">Kelly,

Basham, 1914-1990</persname></origination>
<unittitle label="Title" encodinganalog="245">Basham Kelly papers, <unitdate type="inclusive" normal="1936/1988" encodinganalog="260">1936-1988</unitdate><unitdate type="bulk" normal="1949/1984">1949-1984</unitdate></unittitle>
<physdesc label="Size" encodinganalog="300"><extent>11 linear ft. (25 boxes)</extent></physdesc>
<unitid countrycode="us" repositorycode="xx-x" type="classification" label="Collection No.">MS-F24</unitid>
<repository label="Repository"><corpname>Bluegrass State University. Rodgers Library. University Archives.</corpname></repository>
<abstract label="Abstract">The Basham Kelly papers, 1936-1988, include manuscripts of Kelly's books and articles, personal correspondence with many noted Kentucky writers and musicians, official correspondence from his tenure as chair of the Dept. of English at Bluegrass State University (1949-1984), course material, lecture notes, photographs, and audiotapes and videotapes.</abstract>
</did>
<descgrp>
<head>Administrative Information</head>
<acqinfo encodinganalog="541">
<head>Provenance</head>
 The Basham Kelly papers were donated by Mary Lilly Kelly to the University Archives, Bluegrass State University, in 1991.
</acqinfo>
<accessrestrict encodinganalog="506">
<head>Access</head>
 The collection is open for research use, with the exception of the correspondence files in Series 1, Box 7, which are restricted until 2030.
</accessrestrict>
<userestrict encodinganalog="540">

<head>Publication Rights</head>
 For permission to publish, contact the Curator of the University Archives.
</userestrict>
<prefercite encodinganalog="524">
<head>Preferred Citation</head>
[Item, folder title, box number], Basham Kelly papers, University Archives, Rodgers Library, Bluegrass State University.
</prefercite>
<processinfo encodinganalog="583">
<head>Processing Information</head>
 The collection was processed at the University Archives in 1992 by Judith Morgan. The finding aid was prepared by Diana Elizabeth in 1992.
</processinfo>
</descgrp>
<bioghist encodinganalog="545">
<head>Biographical Note</head>
 Dr. Basham Kelly, who served as the Chair of the Department of English at Bluegrass State University from 1949 until his retirement in 1984, was born in Bullitt County, Kentucky in 1914. He married Mary Lilly, of Georgetown, Kentucky, in 1938. He received his B.A. from Western Kentucky University, his M.A. from the University of Kentucky, and Ph.D. from the University of Iowa. Before joining the faculty of Bluegrass State University, he taught at Stephen F. Austin College and Oklahoma City University.
An influential literary scholar and folklorist, Dr. Kelly was a central figure in Kentucky literary and arts circles for decades, cultivating long-lasting correspondences with numerous novelists, poets, and musicians, including prominent Kentuckians such as Robert Penn Warren, Jesse Stuart, Harriette Arnow, Hollis Summers, Bradley Kincaid, and Bill Monroe.
Dr. Kelly authored four books: Melville’s Politics (1947); Shakespeare in Nineteenth-Century America (1960); Fugitive Traces: Robert Penn Warren and

Contemporary Fiction (1966); and Mountain Music: A Guide to Kentucky Folk Arts (1980). He edited Tall Tales of Madison County (1983), and was a frequent contributor to the Register of the Kentucky Historical Society.
</bioghist>
<scopecontent encodinganalog="520">
<head>Collection Scope and Content</head>
 The Basham Kelly Papers range in date from 1936 to 1968, with the bulk of the material dating from Kelly's tenure as Chair of the Dept. of English at Bluegrass State University (1949-1984). The collection includes: personal correspondence with family, friends, and many notable Kentucky writers and musicians (10 boxes); official English Dept. correspondence (6 boxes); course material, lecture notes, and conference papers (3 boxes); typescript drafts and published editions of all of Kelly's books and articles (4 boxes); six audiotapes and four videotapes of radio and television interviews, lectures and commencement addresses (1 box); and 27 photographs of Kelly and his friends and acquaintances (1 box).
The collection is a valuable primary source for research on Kentucky's literary and folk music scenes, as it contains interesting correspondence from writers such as Robert Penn Warren, Jesse Stuart, Harriette Arnow, and Hollis Summers, and musicians including Bradley Kincaid (the "Kentucky Mountain Boy"), and Bill Monroe, the bluegrass pioneer.
<arrangement>
 The collection is arranged in four series: Personal Correspondence; Official Correspondence, Course Material, Lectures; Manuscripts of Publications; and Photographs, Audiotapes, and Videotapes.
</arrangement>
</scopecontent>
<controlaccess>
<head>Index Terms</head>
 This collection is indexed under the following headings in the online catalog of the Rodgers Library.

<persname encodinganalog="600">Kelly, Basham, 1914-1990.</persname>
<persname encodinganalog="600">Warren, Robert Penn, 1905- </persname>
<persname encodinganalog="600">Stuart, Jesse, 1906-1984. </persname>
<persname encodinganalog="600">Arnow, Harriette Louisa Simpson, 1908-
</persname>
<persname encodinganalog="600">Summers, Hollis Spurgeon, 1916- </persname>
<persname encodinganalog="600">Kincaid, Bradley. </persname>
<persname encodinganalog="600">Monroe, Bill, 1911- </persname>
<corpname encodinganalog="610">Bluegrass State University–Faculty.</corpname>
<corpname encodinganalog="610">Bluegrass State University–Dept. of
English and American Literature. </corpname>
<subject encodinganalog="650">American literature–Kentucky–History and
criticism.</subject>
<subject encodinganalog="650">Folk literature, American–
Kentucky.</subject>
<subject encodinganalog="650">Folk music – Kentucky.</subject>
<subject encodinganalog="650">Folklorists – Kentucky.</subject>
</controlaccess>
<dsc type="combined">
<head>Description of Series/Container List</head>
<c01 level="series">
<head>Series 1</head>
<did>
<unittitle>Personal Correspondence<unitdate type="inclusive"
normal="1936/1988">1936-1988</unitdate></unittitle>
<physdesc><extent>4 linear ft. (10 boxes)</extent></physdesc>
</did>
<scopecontent>
<p>Consists of autograph and typed letters written to Kelly, along
with some copies of letters by Kelly. Includes substantial correspondence from
Robert Penn Warren, Jesse Stuart, Harriette Arnow, Hollis Summers,

Bradley Kincaid, Bill Monroe, and others.
<arrangement>
Arranged alphabetically by correspondent. Letters by Kelly are
filed with letters from correspondents under correspondents' names.
</arrangement>
</scopecontent>
<accessrestrict>
Access to the correspondence files in Series 1, Box 10, is restricted
until 2030.
</accessrestrict>
<c02 level="file">
<did>
<container label="Box" type="box">1</container>
<unittitle>A-D</unittitle>
</did>
</c02>
<c02 level="file">
<did>
<container label="Box" type="box">2</container>
<unittitle>E-G</unittitle>
</did>
</c02> . . . [remaining Series 1 boxes omitted from sample]
</c01>
<c01 level="series">
<head>Series 2</head>
<did>
<unittitle>Official Correspondence, Course Material, Lectures, <unitdate
type="inclusive" normal="1949/1984">1949-1984</unitdate></unittitle>
</did>
<c02 level="subseries">
<head>Subseries 1</head>

<did>
<unittitle>Official Correspondence</unittitle>
<physdesc><extent>2 linear ft. (5 boxes)</extent></physdesc>
</did>
<scopecontent>
Consists of official correspondence written by and to Kelly in his role as Chair of the English Dept. at Bluegrass State University.
<arrangement>
Arranged alphabetically by correspondent or topic.
</arrangement>
</scopecontent>
<c03 level="file">
<did>
<container label="Box" type="box">11</container>
<unittitle>A-G</unittitle>
</did>
</c03>
<c03 level="file">
<did>
<container label="Box" type="box">12</container>
<unittitle>H-J</unittitle>
</did>
</c03>
</c02>
<c02 level="subseries">
<head>Subseries 2</head>
<did>
<unittitle>Course Material, Lectures </unittitle>
<physdesc><extent>1.5 linear ft. (4 boxes)</extent></physdesc>
208 METADATA: A CATALOGER'S PRIMER
</did>

<scopecontent>
Consists of official correspondence written by and to Kelly in his role as Chair of the English Dept. at Bluegrass State University.
<arrangement>
Arranged alphabetically by correspondent or topic.
</arrangement>
</scopecontent>
<c03> [contents of Series 2, Subseries 2 omitted from sample]
</c03>
</c02>
</c01>[contents of Series 3-4 omitted from sample]
</dsc>
</archdesc>
</ead>

Appendix 2:

EAD Standard, The Structural Overview

For the purpose of developing the Method and applying it to *Accademia di Santa Cecilia* EAD encoded metadata it was considered as crucial to include a short structural description of EAD standard. This section is of importance for the reader in order to understand text in Chapters 3 and 4 and occasional references to certain EAD data elements, as some of their semantics can be found in this section.

The EAD DTD is an XML file used to define a set of tags and structural rules for encoding archival finding aids that can be found on-line. EAD DTD allows representation of archival records, which are arranged and described as a hierarchy. There are several different types of hierarchies, but the one used in this case is of *tree* structure, meaning there is exactly **one root**, and each node other than root, has exactly **one parent**, as shown on Figure 1 in Introduction. (Silvello, 2011)

Archivists using EAD will in practice, most likely, be consulting the Encoded Archival Description Tag Library (currently Version 2002) which was also consulted for the purpose of my work. The EAD Tag Library lists all the defined EAD elements (total number of 146) and they can be split evenly into two primary areas:

1. Those containing summary information on the finding aid itself, covered by the Header <eadheader> and Frontmatter <frontmatter> data elements.
2. Those containing summary description of the contents of the archival materials themselves, covered by the Archival Description <archdesc> and Description of Subordinate Components <dsc> data elements. (Pitti, 2005; Carpenter & Park, 2009).

> Furthermore, many elements can be refined with particular attributes, along with the attribute(s)' given value. (For example, Component <c>, can be refined with the attribute @level that signifies the hierarchical level of the archival materials being described (<c level="fond | classification group | group of files | file | item">). (Thurman, 2005)

Some elements can contain text directly, while other elements are intended to help structure the finding aid into sections and therefore must contain other elements inside them. These structural elements are called "wrapper elements" and the outermost wrapper element, used to introduce an encoded archival finding aid, is Encoded Archival Description <ead> data element.

The simplified structure with short description of EAD XML file can be presented as:

<ead>

<eadheader>

> "Fulfills the need for the most basic of publication information access points and general administrative information on the finding aid. It also provides standardization in the inclusion and sequencing of this information across all EAD-encoded finding aids." (Carpenter & Park, 2009, p. 136)

</eadheader>

<archdesc>

> Contains information on the archive itself, and covers a wide range of descriptive information on the context, content, provenance, organization, physical form, and extent of the archival materials, as well as administrative information such as the location of the holding repository, preferred citation

forms, and any restrictions on use or access.
(Carpenter & Park, 2009)
It has two main parts:

<did>

Descriptive Identification element is used to group together key information about the entire body of material being described.

</did>

<dsc>

Description of Subordinate Components is the inventory of repeatable, hierarchically nested **<c>** *Component* tags (e.g., **<c>**, **<**c01**>**, **<**c02**>**) that separates the archival materials into their component units. As shown, attribute @level specifies whether series, subseries, folder, item or other level is being described. Component provides information about the content, context, and extent of a subordinate body of materials. They are usually nested within another <c> data element. Each component may also contain descriptive sub-elements as required. When this descriptive information is included within a component, that information is inherited by all lower level component tags nested within it (Encoded Archival Description Tag Library, 2002; Carpenter & Park, 2009)

</dsc>

</archdesc>

</ead>”

From the above stated, it can be summed up that descriptive information about the archival material is subdivided into different levels of subsets or abstractions. This information, describing particular nodes, can be found within particular elements that are often repeated throughout different levels. Some of these elements are overviewed further in this text.

Description Elements

As it can be noted in the Chapter 3, for the context of this thesis elements containing metadata on the finding aid are not of relevance, and therefore are not going to be discussed further. The information deemed significant is only of some elements, as the reader would get a general idea about the mapping solutions and decisions made in the Mapping tables (Appendices 3 and 4).

Descriptive Identification <did> is a "wrapper" or a data element that groups other elements that are content- based. Those data elements grouped by <did> are mostly optional, and are thought to be among the most important for ensuring a good basic description of an archival unit or component. <did> is used to describe the entire body of material- if it is found at <archdesc> level, or the specific subset of material- if used within Component <c>. However, not every <did> subelement is used at every level of description. At the higher hierarchical levels it usually contains data elements such as: <repository> holding information on the repository where the documents are held , <origination> informing about individuals or organizations responsible for the creation or assembly of the archival materials; <abstract> holding a short abstract of their contents; <unitdate> with creation dates of the archival materials (it is suggested to be normalized by using attribute @normal); <phisdesc> carrying Physical Description; <unittitle> with the title of the unit, dimensions, genre, form. and other physical characteristics. The lower <c> levels are more likely to contain information on the number of the carton, box, folder, or other holding unit in which the archival materials are arranged and stored

(<container>); general comments, citations, or annotations (<note>); links to digital surrogates of the material being described in the finding aid (<dao> and <daogrp>), and so on. Other elements that are mainly used for enriching contextualization, usually used at the higher levels of hierarchy are: Biography or History <bioghist>, which could contain a concise essay about the life of an individual or family or about the history of a corporate body and Scope and Content <scopecontent> statement, a prose overview of the topical content. (Ruth, 2001)

To provide access to the materials through authority controlled searching, the Controlled Access Headings (<controlaccess>) wrapper data element is used to group access points that can be: personal name <persname>, corporate name <corpname>, geographical name <geogname> , genre <genreform>, subject <subject> or title <title>.

For the purpose of linking to the sources, both in and out of the finding aids, several elements exist: <ref>, <archref>, <bibref>, <ptr>. For the assigning electronic representations of the materials being described Digital Archival Object <dao> data element is used.(Thurman, 2005). Mark-up tags that can be used in order to format the text are: <head>, <p>, <list>, <abbr>,<table>, <emph> and other.

This chapter gave a glimpse into the complex structure of EAD .

In conclusion, the archival context of the tree structure is commonly instantiated by means of a unique XML file (see Appendix 1 for an example) which mixes up the hierarchical structure elements with the content elements, without a clear distinction between the two. This makes it not straightforward to determine how to access and exchange a specific subset of data without navigating the whole hierarchy or without losing meaningful hierarchical relationships. (Silvello, 2011) Chapters 3 and 4 are going to deal with this issue through transforming the information from EAD to EDM.

Appendix 3

Europeana Data Model (EDM)

This section looks carefully at the goal language used in this thesis for remodeling archival data encoded in EAD schema. It will present the most important features and the structure of Europeana Data Model, background of which can be seen in the Chapter 1.
It should be noted this model is relatively recently published and there still has not been much publications covering it. For this reason, main source of information is obtained from only three documents: Europeana Data Model Primer (2010), Definition of Europeana Data model Elements (2011) and The Europeana Data Model (Doerr at al., 2010).

Vision behind creating the model

As stated by Concordia at all.(2009) the vision of Europeana is not to be merely a portal but rather to exploit the great amount of data aggregated from the different cultural heritage institutions and offer it to all sorts of external communities (such as eScholarship collaboration or various digital libraries) to reuse for their own needs by means of API. However, the idea behind this vision is to go even one step further from the traditional digital library interface and to offer rich semantic contextualization for the object representations in Europeana in a way that would enable complex semantic operations on these resources. The way this was intended to achieve is by interconnecting surrogate objects that represent born digital or digitized cultural heritage object provided to Europeana, as the first abstract layer, and additionally contextualizing them with links to nodes of a semantic network, forming the second data layer in Europeana. This view is illustrated in Figure 3.

In this figure (Fig. 3), the blue circles on the lower level show constituents of a Digital Surrogate Object, such as related metadata, licensing information, abstractions annotations and their representations. At the same time those objects have contextual links to other objects as well as to concept nodes (purple circles) such as those representing time and space entities or abstract concepts (so called non information resources).(Concordia at al., 2009)

In order to offer this rich semantic contextualization for the object representations, they need to be systematically connected to Linked Open Data (LOD) project (http://linkeddata.org/) on the WWW or to semantic contextualization resources held within the Europeana data space, such as thesauri and structured vocabularies migrated to the SKOS standard. (Doerr at al., 2010)

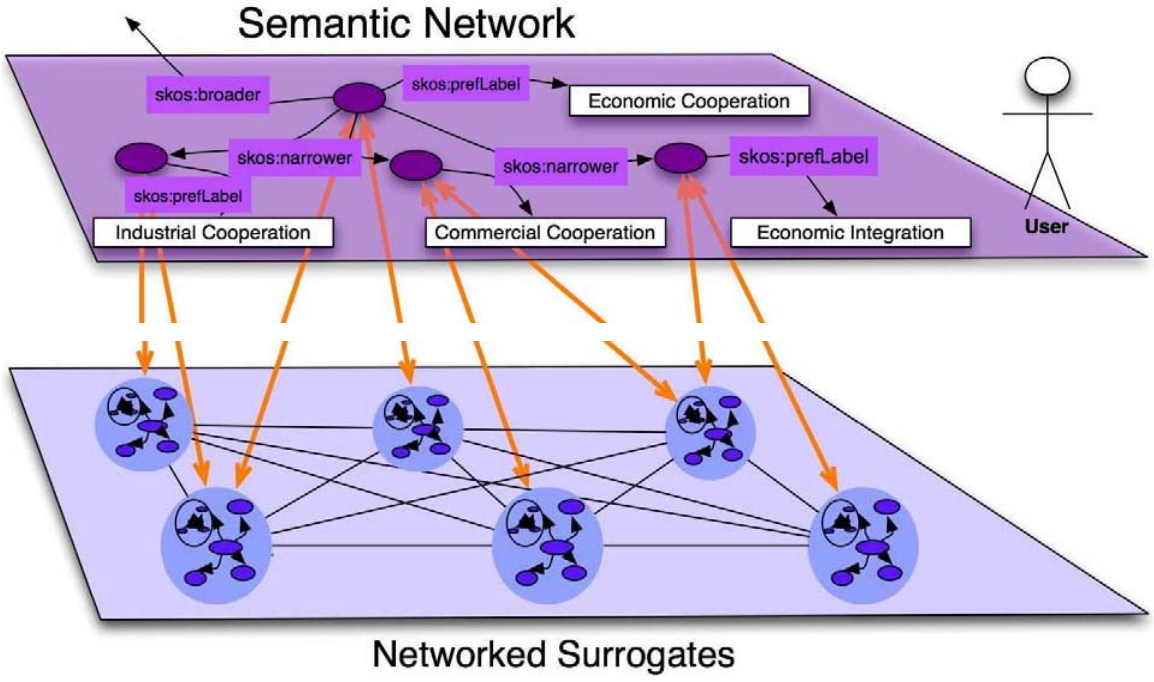

Figure 3: Semantic Network and Networked Surrogates in Europeana

EDM Model

In order to support the previously stated vision, a new data model, Europeana Data Model (EDM), was developed and its design was informed to a large extent by the vision of Semantic Web. The Semantic Web, as articulated by Tim Berners-Lee (1999), is the vision of a Web in which resources are accessible not only to humans, but also to automated processes, e.g., automated \agents roaming the Web performing useful tasks such as improved search (in terms of precision) and resource discovery, and so on.

Therefore, the building blocks of EDM can be presented as the modified picture of Semantic Web "layer cake". This modified "layer cake" can be seen in the Figure 4 and the concepts underlying its building blocks, are to be discussed to some extent in further text.

XML is a metalanguage allowing users to define markup for their documents using tags, and it is pervasively used as the data encoding and interchange standard in cultural heritage institutions. However, it does not provide any means of talking about the semantics (i.e. meaning) of data. Therefore, for the basis of EDM data model RDF (Resource Description Framework) is used. The rational behind RDF is that resources can be described by means of semantically meaningful connections between them, as it allows representing structured information about any resource in the form of a simple triple statement (subject, predicate, object). (Doerr at al., 2010).

Except for the statement, other two fundamental concepts of RDF are resources and properties. Resources are the "things" we want to talk about, which can be subjects, predicates or objects in statements. For indentifying such resources RDF uses URIs (Uniform Resource Identifiers) as the basis of its mechanism, "All URIs share the property that different persons or organizations can independently create them, and use them to identify things." (RDF Primer, 2004, chapter 2.1 para. 17) Properties are a

special kind of resources, which describe relations between resources, and they can be of different type e.g.: “written by”, “age”, “title” etc. (Antoniou & van Harmelen, 2008)

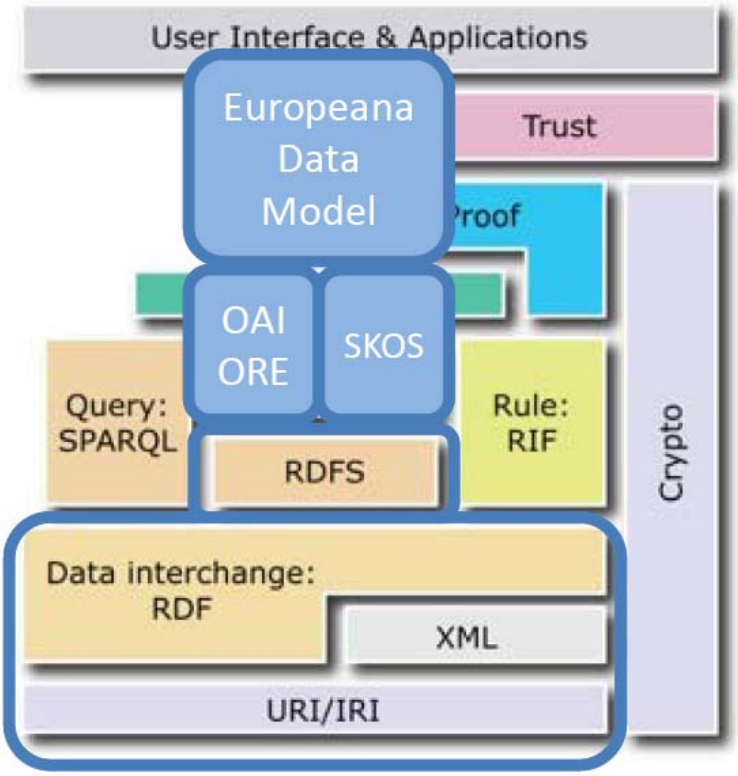

Figure 4: Europeana Data Model “Layer Cake” (Meghini, n.d , p.3)

The example of RDF triple can be: (ec:ulysses, ex:author, ex:james_joyce) describing the book Ulysses by connecting its identifier (URI) to another that stands for James Joyce, using an author typed property which denotes the relation between a book and its author. (Doerr at al., 2010) While properties are always presented by URIs, a node in an RDF graph may be either a URI, a literal value, or blank (having no form of identification independent of the local graph). Since the object of one statement may be the subject of another, sets of statements may also be considered as graph structures, in which the

subjects and predicates appear as nodes linked by (property) arcs (Styles, Ayers, & Shabir, 2008). The example RDF graph can be seen in Figure 5:

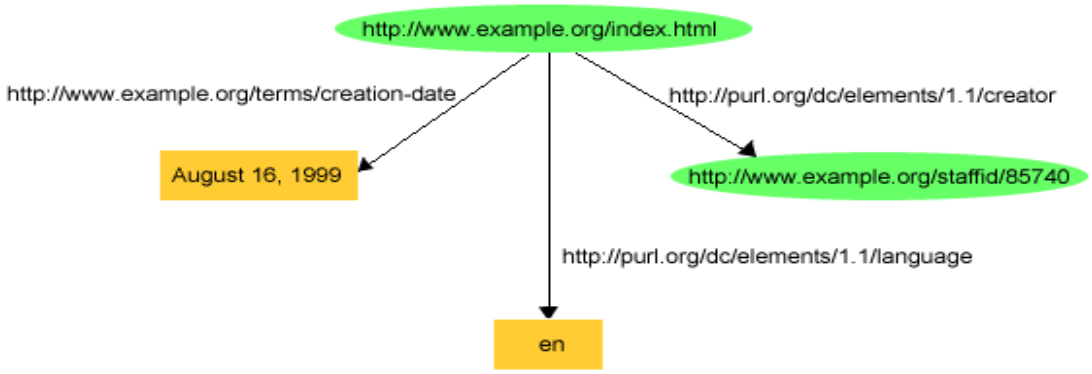

Figure 5: Several RDF Statements About the Same Resource (RDF Primer, 2004; chapter 2.2)

Types can be assigned to the resources, and for the subjects and objects this is done by making them instances of a particular class defining their type. Furthermore, for the classes of particular domain, constrains and rules are declared on the possible relations between them. These rules defined by ontologies. The most used definition of ontology is the one by Gruber (1993) "a specification of a conceptualization" and they are defined by means of the RDF Schema (RDFS) and Web Ontology Language (OWL) standards. This mechanism was chosen for specifying the domain covered by Europeana, since "running an inference engine on top of data for a collection and books and paintings, and querying for all objects created by one person would allow retrieving all these objects without prior knowledge of their specific type, a crucial feature when information integration is required". (Doerr at al., 2010; p. 3; Synak, Dabrovski, Kruk, 2009)

EDM reuses some of the reference ontologies already available. One is the W3C standard Simple Knowledge Organization System (SKOS), which defines a model to represent the elements of Knowledge Organization Systems (KOS) such as thesauri, classification schemas, taxonomies and their likes in a machine readable way by means of RDF. It has been used by many initiatives, some of which from the library domain are Library of Congress that published its Subject Headings (LCSH) in SKOS

(http://id.loc.gov/authorities/about.html) and Universal Decimal Classification in SKOS (http://www.w3.org/2006/07/SWD/wiki/EucUDC). SKOS features a main class to describe concepts, by providing for their lexical properties (skos:prefLabel and skos:altLabel), further it allows expression of semantic relations between them (skos:narrower, skos:broader, skos:related) and documenting them by adding notes and description (e.g. skos:scopeNote, skos:definition). An instance of previously mentioned relationships is illustrated in Figure 6, that demonstrates how thesauri can be presented in SKOS.

Another important functionality SKOS allows is matching across concept schemas by, for example, linking concepts from different thesauri, that are semantically equivalent, using the skos:exactMatch property. Europeana uses SKOS in order to build the semantic data layer, i.e. creating a layer of interconnected controlled vocabularies that can be used to enrich existing object metadata. (Olensky, 2010) This enables applications to navigate through a semantic layer of concepts from different sources, leveraging such conceptual network to access objects that are originally described using different, but semantically related concepts as shown in Figure 3. (Doerr at al., 2010)

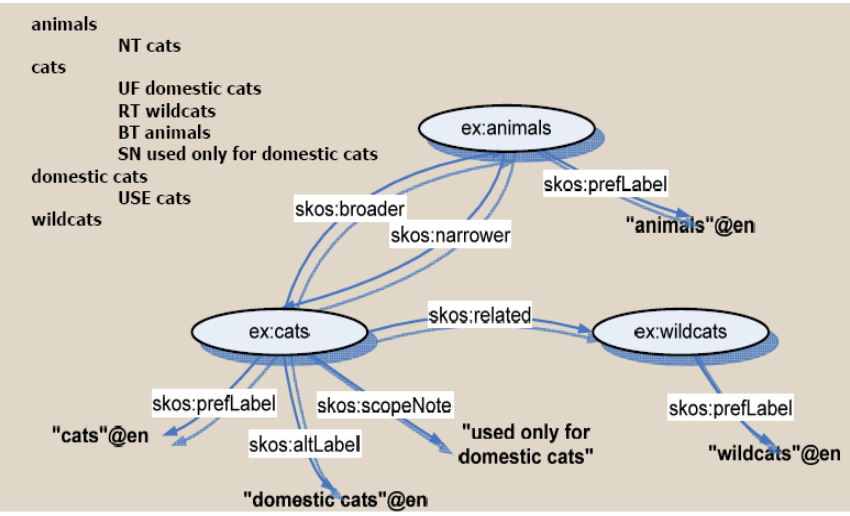

Figure 6: Thesauri presented by means of SKOS (Isaak, n.d. p.17)

Another ontology included in EDM is Dublin Core (http://dublincore.org/)which is among the most famous and widely used metadata element set in the Digital Library domain. DC was established by an international, cross disciplinary group of professionals, and is maintained by open organization called Dublin Core Metadata Initiative (DCMI). “Dublin Core gives a compact vocabulary to describe the core features of culture objects (creators, relations to other resources, subject indexing, etc.) in a Semantic Web-enabled fashion that fits a very wide range of needs.” (Doerr at al., 2010; p.4) This allows institutional providers who already have their data encoded in DC to keep to a simple vehicle for providing their data, but also supports sharing and re-use between EDM data and other applications running on this standard. Furthermore, it allows the legacy Europeana data already encoded in ESE to be injected in the new model. (Doerr at al., 2010)

There is a well-indentified set of elements EDM uses to carry out its task, and those reused from other schemas, are:

• The Resource Description Framework (RDF) and the RDF Schema (RDFS) namespaces (http://www.w3.org/TR/rdf-concepts/)
• The Simple Knowledge Organization System (SKOS) namespace (http://www.w3.org/TR/skos-reference/)
• The Dublin Core namespaces for elements (http://purl.org/dc/elements/1.1/, abbreviated as DC), terms (http://purl.org/dc/terms/, abbreviated as DCTERMS) and types (http://purl.org/dc/dcmitype/, abbreviated as DCMITYPE)
• The OAI Object Reuse and Exchange (ORE) namespace (http://www.openarchives.org/ore) (primer)

Furthermore. EDM introduces its own set of elements, for which Europeana Namespace is used. As of February 2011, the RDF schema for the namespace http://www.europeana.eu/schemas/edm/ is not yet in place. (Definition of the Europeana Data Model Elements, 2001) For the sake of describing classes and properties in this

chapter (as in chapters 3 and 4) namespace “ens:” is going to be used, as it was used in the Europeana Data Model Primer.

The EDM class taxonomy can be seen in the following Figure 7, where the classes introduced by EDM are shown is light blue rectangles. The classes in the white rectangles are re-used from other schemas.

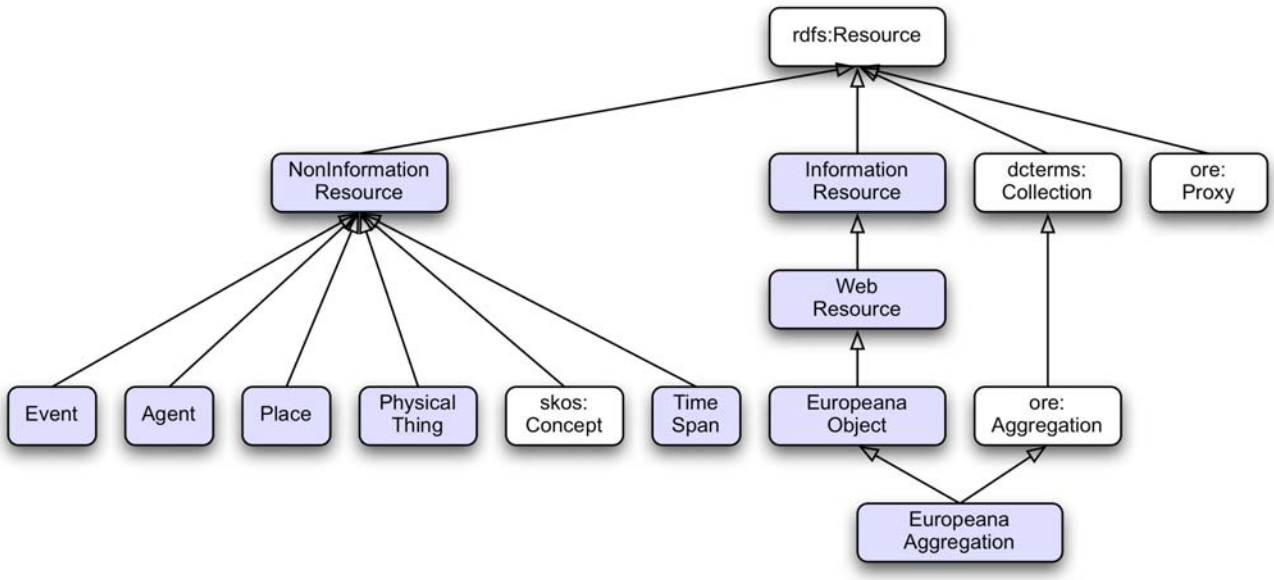

Figure 7: The EDM Class hierarchy (Definition of the Europeana Data Model elements, 2001, p. 5)

The EDM property hierarchy, without the properties included in ESE is represented below in Figure 8. The majority of the properties in this figure are defined by Europeana namespace (in light blue rectangular), while the classes in the white rectangles are re-used from other schemas. Still, the area where further research is necessary is on the reuse of properties from existing metadata schemas or ontologies (e.g. the EDM properties ens:wasPresentAt, ens:happenedAt and ens:occurredAt are directly taken from the CIDOC CRM ontology, yet not identified as such. (Olensky, 2010) The other properties used from DC schema are not included in the figure.

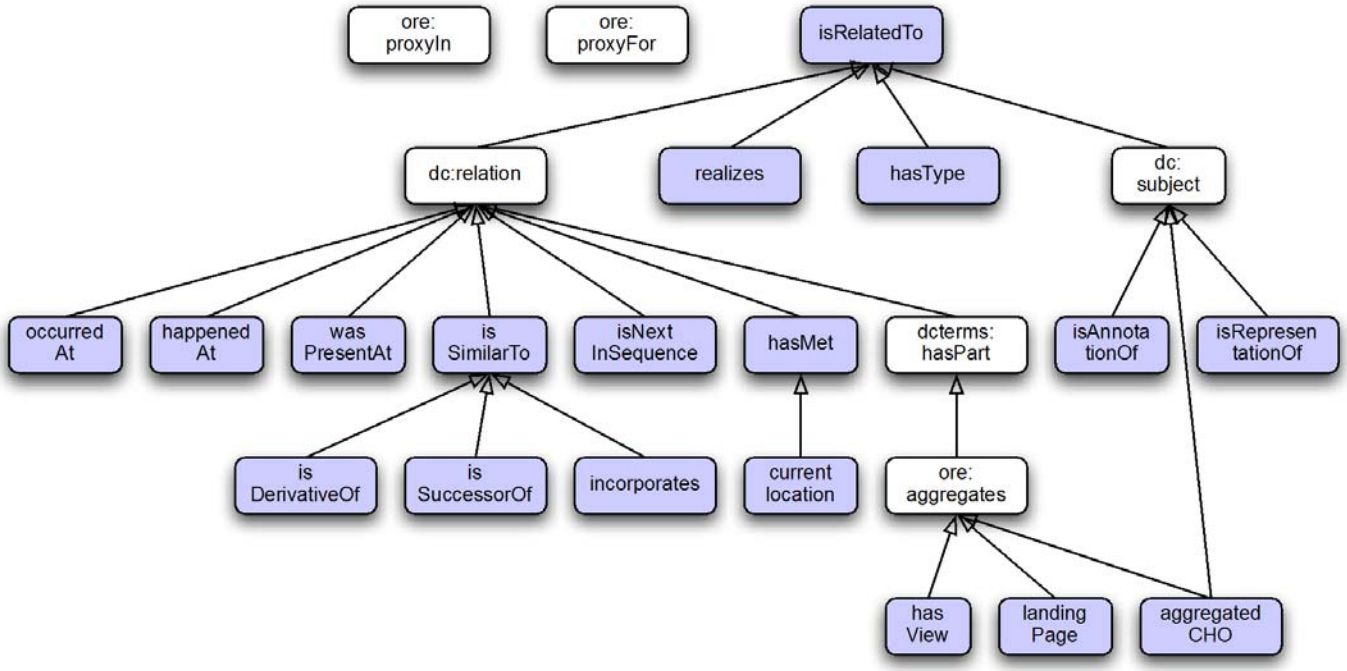

Figure 8: The EDM property hierarchy without the properties included in ESE (for readability). (Definition of the Europeana Data Model elements, 2011, p. 14)

The Way of Representing Data in EDM Using OAI-ORE

In EDM for the structural modeling framework OAI Object Reuse & Exchange (OAI-ORE) specifications (http://www.openarchives.org/ore/1.0/to) were chosen. OAI-ORE is maintained by the Open Archives Initiative, which develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content. OAI-ORE is also based on and influenced by RDF model. Moreover OAI-ORE advocates use of recent developments in the areas of the Semantic Web, Linked Open Data and Cool URIs. (Tarrant, O'Steen, Brody, Hitchcock, Jefferies & Carr, 2009, Open Archives Initiative, n.d.) The basic idea behind OAI-ORE are concepts of Aggregations and Aggregated Resources, where an Aggregation is simply a set of Aggregated Resources, all of which are represented by URIs. Figure 9 illustrates an example of a publication as an OAI-ORE Aggregation.

While the Figure 9 serves as an example of only a single publication record as an Aggregation is demonstrated, the abstract concept introduced by OAI-ORE also allows nesting of Aggregations. “For example, the highest level Aggregation could be the repository and the Aggregated Resources thus become the publications, which in turn contain their own Aggregated Resources.” (Tarrant at al., 2009, para. 21) As it can be seen in Chapter 3, this mechanism is used for representing archival finding aids. While there is no limit to the depth with which resources can be aggregated, it is not recommended to go to too many levels due to the recursive operations that will need to be performed to import these resources. (Tarrant at all, 2009).

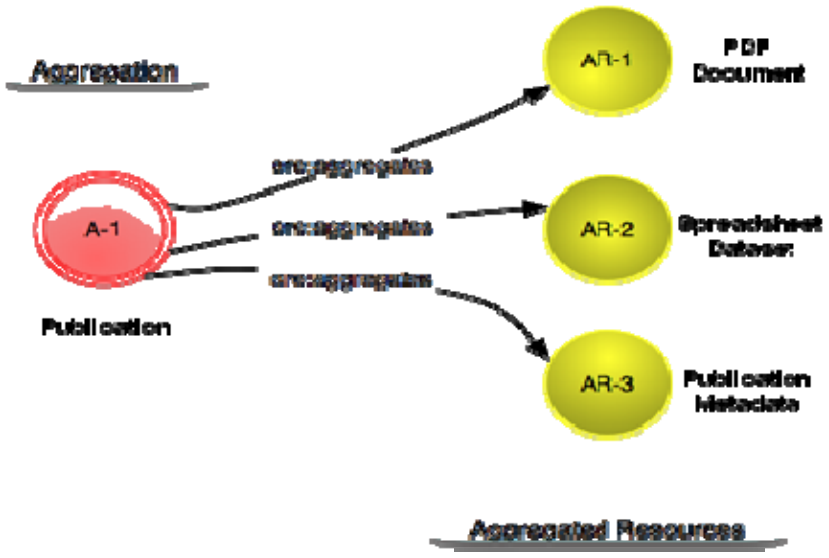

Figure 9: Example OAI-ORE Aggregation of a Publication. (Tattant at al., 2009, para. 20)

.

EDM considers two basic classes of resources provided to Europeana:
- the “provided object” itself (e.g. painting, movie, music score, book) and
- a (set of) accessible digital representation(s) of this object, some of which will be used as previews (e.g., a thumbnail of a painting’s digital picture).

Together they form one Aggregation. To relate Aggregation to its Aggregated resources, EDM is using two properties: ens:aggregatedCHO, where CHO stands for Cultural Heritage Object, and ens:hasView property, for one or more resources that are digital representations of the provided object. How this mechanism works is demonstrated in the Figure 10. This feature allows capturing the distinction between "works", which are expected to be the focus of users' interest as described by metadata in the records, and on the other hand their digital representations, which are the elements manipulated in information systems like Europeana. (Europeana Data Model Primer, 2010)

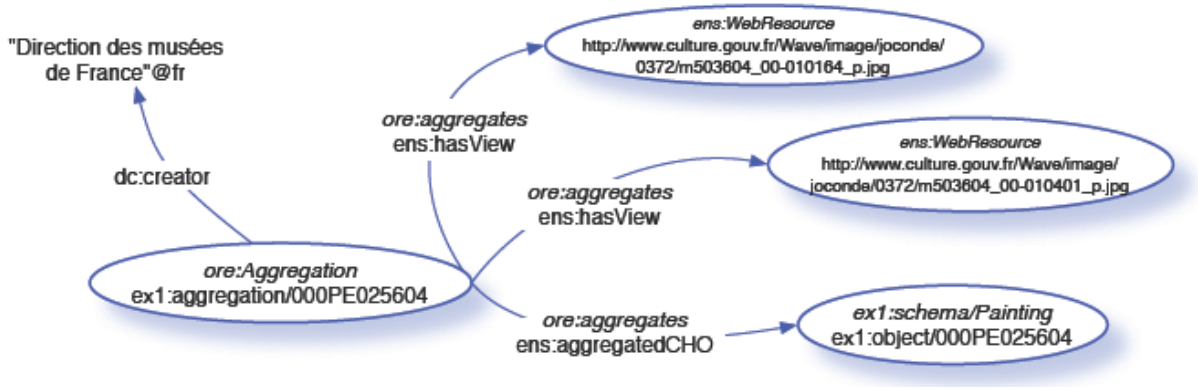

Figure 10: One provider's aggregation and provided object (Europeana Data Model Primer, 2010, p. 11)

One of the mechanisms from OAI-ORE, which is providing the key functionality to the EDM model is the use of Proxies (http://www.openarchives.org/ore/1.0/datamodel#Proxy). Proxies are used to enable the representation of different views on the same resource. The rational behind using Proxies is that several institutions may provide different "views" on the same resource, e.g. different names may be used for the same creator. Furthermore, Europeana will attempt to add its own enriched data about that resource giving yet another view on the same resource. In the future, the user may add to this by giving his/her particular view of the resource. Proxy mechanism allows keeping all the different views separated, so the provenance of the description is easily referable. Therefore, an aggregation can be seen as one provider's contribution for an object, or to say one "view" of the object, it can give

raise to only one Proxy per object that it aggregates. In other words, Proxy will hold the descriptive metadata on the provided object from one contributor. (Europeana Data Model Primer, 2010)

The example from the Europeana Data Model Primer, slightly modified (Fig. 11), shows how Proxies are used in the case where there are two providers for one Cultural Heritage Object, which is this case painting of Mona Lisa as seen from 2 institutional providers of description. Figure 11 demonstrates: providers aggregations (ore:Aggregation ex1 and ex2), provided object (connected by ens:aggregatedCHO property to Aggregations) and Proxies (ore:Proxy ex1 and ex2) together with the description they carry (outlined in red).

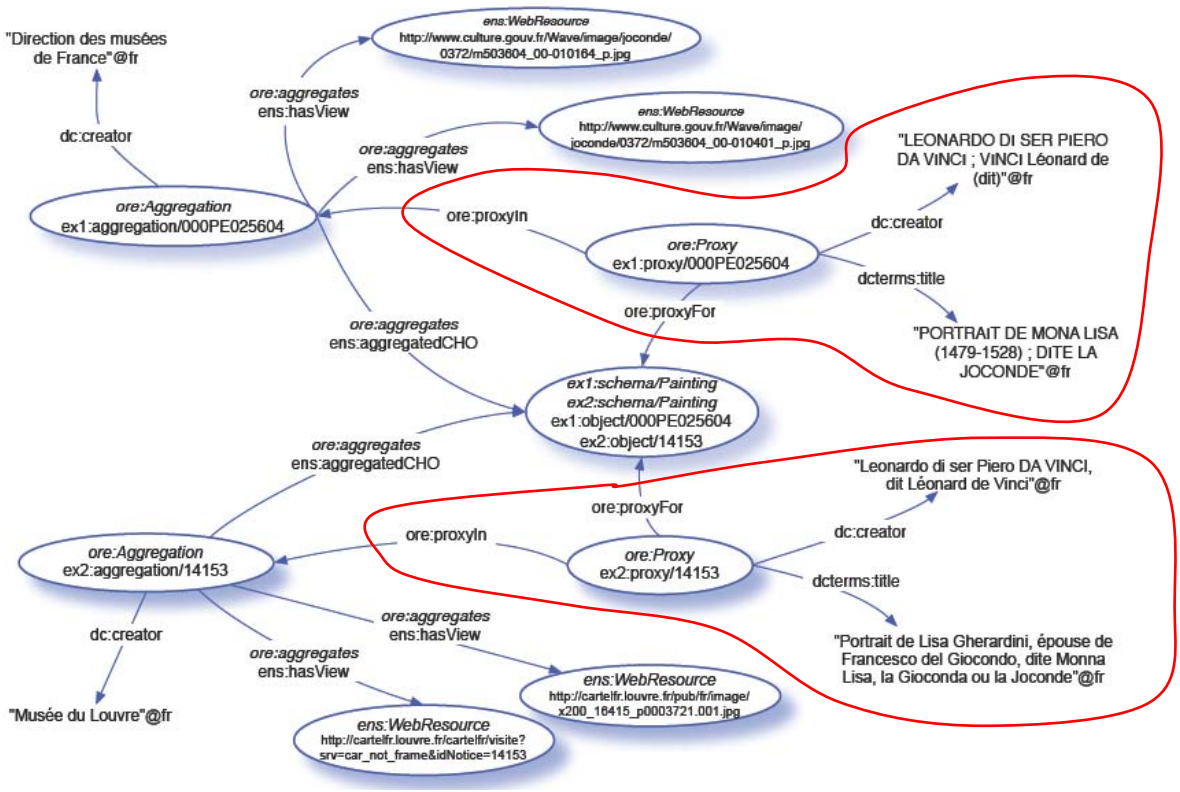

Figure 11: Providers' aggregations, provided object and proxies—complex case with two providers for the object (Europeana Data Model Primer, 2010, p.14)

Presenting Metadata in EDM

The complexity of EDM, shown in classes and properties presented in taxonomy figures (Fig. 7 and Fig 8) allow both "object-centric" and "event-centric" approach for describing cultural heritage objects.

On one hand, object-centric approach focuses on the object described and information comes in the form of statements providing a direct linking between the described object and its features. Behind this approach stands a century old tradition of librarianship. An application of such approach can be found in Dublin Core metadata set. (Europeana Data Model Primer, 2010)

The Figure 12 shows how metadata is presented in the object-centric approach:

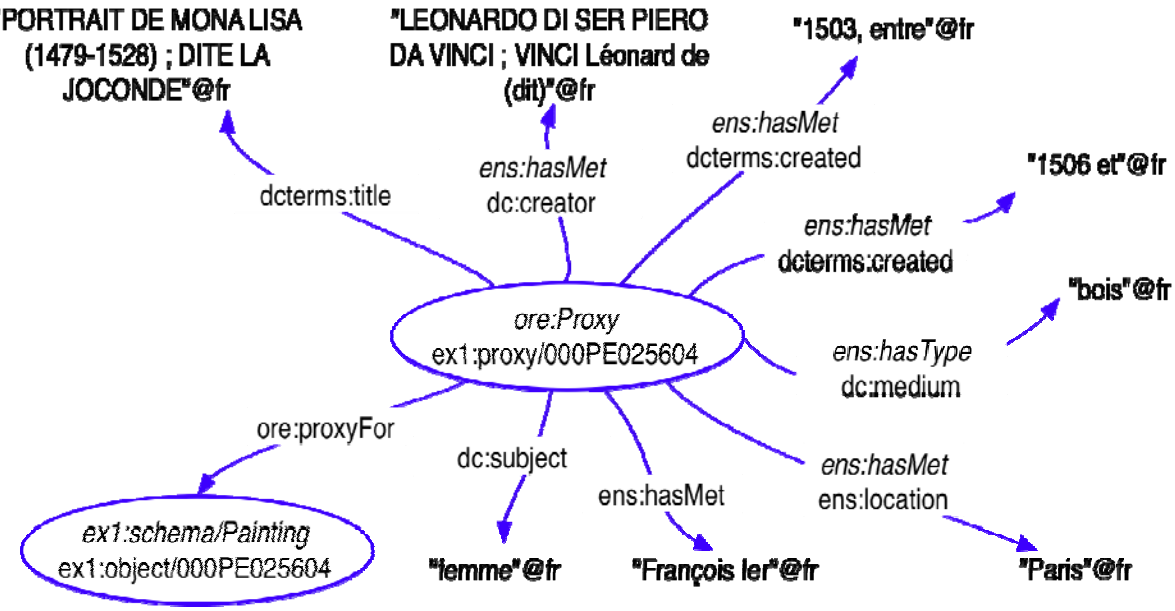

Figure 12: Mona Lisa – an object-centric description (Europeana Data Model Primer, 2010, p. 19)

On the other hand, for the purpose of describing "non-verbose" objects, such as images and objects in museum, event-centric approach is more suitable, as it considers that descriptions of objects should focus on characterizing the various events in which objects have been involved (i.e. archeological finding: excavation, deposition, production etc.). This allows to provide more expressive and coherent records of the provenance and histories of objects. In addition to providing more details on the object, this approach allows for detecting with high precision objects related through a common history, which is synonymous to shared participation in events. (Doerr at al., 2010) This method underlies models mainly used in museum sphere such CIDOC-CRM and LIDO. In EDM events are introduced to the object's described using the class ens:Event.

These events act as the "hubs" relating the object to other entities that were directly connected to it. Relations can be represented in EDM using tree following properties:

- ens:wasPresentAt, between any resource and an event it is involved in;
- ens:happenedAt, between an event and a place;
- ens:occurredAt, between events and the time spans during which they occurred.

(Europeana Data Model Primer, 2010)

How the same information shown in the Figure 12 can be expressed through event-centric approach, as illustrated in Figure 13.

EDM allows both object-centric and event-centric approaches to co-exist seamlessly for the same object. However, EDM properties and classes have much more developed object-centric "core" based on Dublin Core elements than the event-centric (the previously mentioned 3 properties and one class). The authors of the model justify this decision based on the facts that this approach is much more widespread and Dublin Core is a simple, commonly used standard. (Europeana Data Model Primer, 2010)

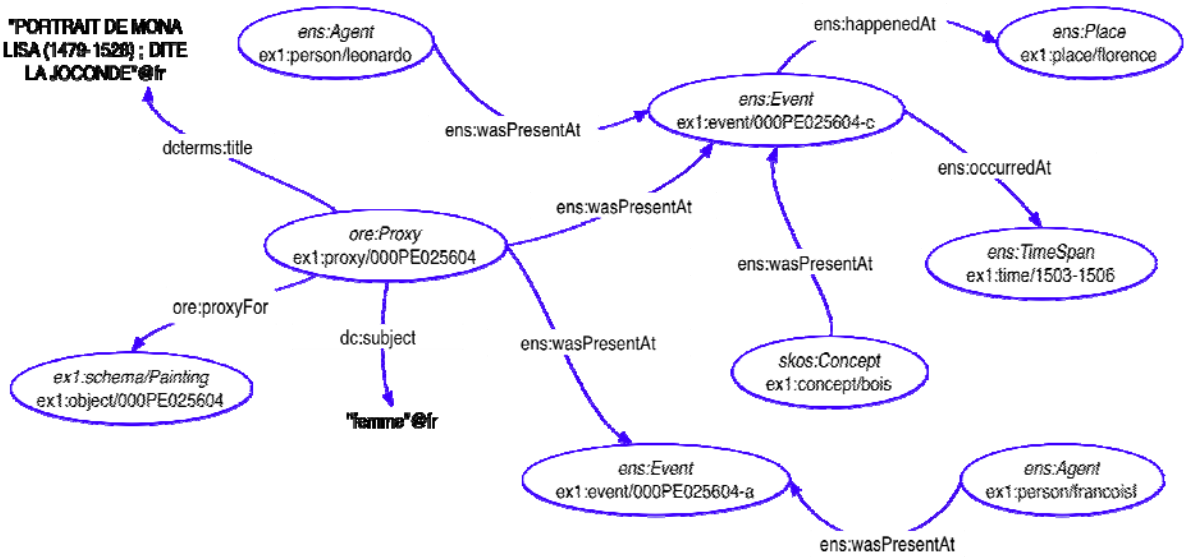

Figure13 Mona Lisa – an event-centric description

Expressive power of EDM

Doerr at al. (2010) , distinguish five fundamental semantic relationships that EDM model offers:

1)Classification into categories expressed with SKOS (by using class skos:Concept).

2)Part decomposition of anything and incorporation of information resources into another one, that allows presentation of complex or hierarchical objects (e.g. using properties dcterms:HasParts and dcterms:isPartOf)

3)Similarity, i.e. the relation between things or information resources sharing some common features by chance, by influence or by a related derivation history as described by FRBR (Doerr and LeBoeuf, 2007). The most general property to express relation are ens:isRelatedTo and its specialization ens:HasType. Different types of derivation can be expressed further, for example by properties such as ens:isDerivativeOf,

ens:isSucessorOf, ens:Incorporates. Properties that may express different types of versions can be ens:isRepresentationOf or ens:Realizes.

4) Aboutness, i.e., the entities or ideas a thing or information resource represents, presents, refers to or is about.

5) History of an item, i.e., the things, people, places, times, events something had contact with, has existed at, "has met" (for example by using the relationship ens:HasMet that provides wide semantic coverage). All historical relationships can be explained and expanded as presence in events and the related event parameters. (Doerr at al., 2010, Definition of the Europeana Data Model elements, 2011)

Those relationships, can be further specialized, for instance by Dublin Core. However, it is expected that providers will submit the data to Europeana that fits their own specific levels of interest. The specific data needs to be mapped to appropriate EDM classes and properties that present a semantic interoperability core. The EDM properties are required to ensure that at least a part of the intended semantics for the specific properties that are exploited. Such mappings are typically achieved in RDF by asserting semantic relationships between the specific constructs and the core ones by taking the form of statements using rdfs:subClassOf or rdfs:subPropertyOf. (Europeana Data Model Primer, 2010)

"This co-existence between the generic and the specific level allows for example:

- -to search for the painting using a generic description-based index

  -to display the information for that painting using the finer-grained distinctions made by the provider."

(Europeana Data Model Primer, 2010, p. 23-24).

The validation attempts were performed at the community meetings, where model was discussed and mapping cases were done from different community standards. The community standards on which the validation was performed are museumdat and LIDO

(from museums sector), MARC (from libraries sector) and EAD (from Archives sector) These validation attempts demonstrated that the EDM has the potential to successfully function as a common top-level ontology for many different kinds of more specialized data models from a various knowledge domains. This is because the model is amenable to declaration of domain specific Application Profiles in order to enrich the precision of EDM for their particular subset of data. (Heery and Patel 2000, Doerr at al., 2010) For example, suggestions from the library experts was that the introduction of RDA (Resource Description and Access) would substantially intensify the need to include the FRBR categories, eventually, as a part of Library community application profile. (Minutes: Mapping Librarian Data to the EDM, 2010)

Normalization

The data submitted to Europeana and conformed to the EDM model is planned to be further normalized by the internal activities in order to reach the vision shown in Figure 3. Current activities in this field, particularly at the VUA (Vrije Universiteit Amsterdam) and the NTUA (National Technical University of Athens), include research on methods for aligning as well as mapping vocabularies, (semi-) automating these workflows, thus reducing the human intelligence factor and methods for fuzzy matching. (Olenskly, 2010) The idea is also to enrich the provided information by giving it more context, by exploiting the rich structural data that can be found in LOD Cloud. “Linked Data adds a fundamental dimension to this vision, because through Linked Data Europeana can use the HTTP URIs in its information space also as links enabling access to structured descriptions of the corresponding objects. These links act therefore as connectors of the Europeana information space with the information space of other authorities, allowing Europeana to collect additional knowledge about people, places, concepts, and so on. Needless to say, the so collected knowledge is expected to play a major role for improving the usability of Europeana in important aspects such as the performance of the discovery functionality…” (Doerr at al., 2010, p.6)

Apendix 4:

One “branch” of Ethnomusicology fond EAD XML
(as found in original data, without <processinfo> data element)

```
<?xml version="1.0" encoding="windows-1252"?>
<dsc>
<c level="fonds" id="ANSC00000001" audience="internal">
  <did>
    <unittitle>Archivio di Etnomusicologia</unittitle>
    <unitid countrycode="IT"repositorycode="ANSC">ANSC.00001</unitid>
    <repository>Archivio etnomusicologico dell'Accademia di Santa
Cecilia</repository>
  </did>
  <dsc>
        <c audience="external" id="ANSC00000002" level="recordgrp">
        <controlaccess>
          <geogname role="regione">Sicilia</geogname>
          <geogname role="stato">Italia</geogname>
        </controlaccess>
        <did>
          <container type="raccolta">001</container>
          <unitid countrycode="IT"
repositorycode="ANSC">ANSC.00001.00001</unitid>
          <unittitle>
            <bibseries>
        <ptr target="00006448" title="Nataletti, Giorgio"/>Giorgio  Nataletti
            </bibseries>
            <num type="raccolta">001</num>
            <ref target="ASC0000002028"/>Giorgio Nataletti
            <unitdate>2.8.1948</unitdate>
```

```
        </unittitle>
      </did>
      <dsc>
       <c audience="internal" id="ANSC00000003" level="item">
         <controlaccess>
              <geogname role="località">Catania </geogname>
              <geogname role="regione">Sicilia</geogname>
              <geogname role="stato">Italia</geogname>
              <persname role="esecuzione" rules="voce maschile,
scacciapensieri">Turi Pandolfini</persname>
         </controlaccess>
         <descgrp encodinganalog="ISAD 5 Allied materials area">
              <originalsloc type="supporti">
                <list>
                     <item>DAT racc. 1 - 2</item>
                     <item>traccia 1</item>
                     <item>Bob. or. RAI-15-175936-7</item>
                     <item>Copia RAI-15-175936-7</item>
                </list>
              </originalsloc>
         <relatedmaterial type="allegati">
                <list>
                  <item>Schede CNSMP con trascrizione dell'incipit musicale</item>
                  <item>Fogli di registrazione RAI</item>
                  <item>bozze per etichette discografiche</item>
                                                          </list>
         </relatedmaterial>
              </descgrp>
      <did>
         <container type="raccolta">001</container>
         <materialspec type="durata">1' 38"</materialspec>
```

<unitid countrycode="IT" repositorycode="ANSC">
ANSC.00001.00001.00001</unitid>
<unittitle>
<bibseries>Giorgio Nataletti</bibseries>
<geogname>Catania (ma a Roma) - studio RAI</geogname>
<num>001</num> Aria
<unitdate normal="19480802 19480802">2.8.1948</unitdate>
</unittitle>
</did>
<note encodinganalog="ISAD 6 - 1 note">
Nella pubblicazione "La ricerca e lo studio dei linguaggi musicali della Sicilia dal 1948 al 1969 attraverso l'opera del CNSMP", curata dal Centro Nazionale Studi di Musica Popolare, Nataletti si sofferma ad illustrare i rapporti di conoscenza con Turi Pandolfini e le modalità di registrazione della raccolta. Dalla lettura di tale testo sembra di capire che i primi due brani dovrebbero essere relativi alla provincia di Catania, mentre il terzo farebbe parte del repertorio di Siracusa. La denominazione locale dello scacciapensieri è maranzanu.
</note>
<odd type="fonte_titolo">
schede CNSMP
</odd>
<metadigit>
<audio>
<sequence_number>1</sequence_number>
<nomenclature>Aria</nomenclature>
<proxies>
<usage>conservativo</usage>
<usage>external</usage>
<file href="/ASC/ETN/000/000/03/ASC.ETN.00000003.0001.mp3"/>
<audio_dimension>
<duration>1' 38"</duration>

```
        </audio_dimension>
        <audio_metrics>
          <samplingfrequency>44.1 KHz</samplingfrequency>
          <bitrate>128</bitrate>
        </audio_metrics>
        <format>
         <name>MP3</name>
         <mime>AUDIO/MP3</mime>
         <channel_configuration>2 ch</channel_configuration>
        </format>
</proxies>
</audio>
</metadigit>
</c>
```

Mapping Table 1; ENS FOND (Fond "Ethnomusicology")

Legend:
- (round brackets) hold the domain of the property, default property is ore:Poxy of the node;
- "/" no mapping;
- "~" same mapping as for the same element above, in these cases only the main elements are displayed without their sub-elements and;
- "->" look right;
- "&" see example 2, p...;
- "*" see example 1, p. ...;
- -"^" see example 3, p. ..
- -"#" attribute

	a) ANSC EAD ELEMENTS	b) EXAMPLES OF SEMANTIC OF ANSC EAD ELEMENTS	c) FIXED VALUE OF THE ATT.	d) EDM MAPPING ELEMENTS	e) ADDITIONAL WORK ON MAPPING
1	<c>	fond , highest node		create instance of ens:ArchivalFond, domain of: ens:IsPartOf (to recordgrp Proxies)	create subclass of ens:NonInformationResource called ens:ArchivalFond
2	#level	fond	fond	dc:type	
3	#id	identifier		dc:identifier	
4	#audience		internal	/	
5	<did>			/	
6	<unittitle>	Archivio di Etnomusicologia	dc:title	dc:title	
7	<uitid>	call number/reference code, value not mapped		/ ens:currentLocation	create instance 1 of ens:Place
8	#countrycode	IT		ens:country (to 1:Place)	
9	#repositorycode	ANSC		dc:source (to 1:Place)	instance 1 of class:Agent, skosaltlable:ANSC, this URI will hold all the data on ANSC, address..+ skos altlable ANSC (to 1:Agent)
10	<repository>	Archivio etnomusicologico dell'Accademia di Santa Cecilia		dc:Alternative (to 1:Place)	
11	<dsc>	wrapper		/	
12					
13	<c>	recordgrp/ item		ore:Aggregation	for every recorgrp Proxy, dcterms:isPartof is going to be declared for the above!! that will hold info on <c level "fond"> highest node
14	#level		recordgrp/ item	dc:type	
	#id	identifier		dc:identifier	
16	#audience			/	
17	<did>			/	
18	<container>	sequential number of container		/	information repeated in <unititle>
19	#type …	collection, box	raccolta	/	
20	<unitid>			~	
21	<unittitle>	title		dc:title	create 1:BN
22	<bibseries>	information about series, #PCDATA		dc:source (to 1:BN)	create 2:BN
23	<ptr>			/	
24	#target	target is authfilenum of the person who gathered this coll.		dc:identifier (to 2:BN)	
25	#title	usually the name of the collector who is the source		dc:title (to 2:BN)	
26	<num>	number of collection or recording		dcterms:alternative	create (3:BN), indicator for making ens:NextinSequence with the previous element at the same level, if the number value is sequential. If the number is the same, sequence should be made after the alphabet order of the letters following the number
27	#type	raccolta-for recrdgrp, brano -for item	raccolta/brano	->	**for the recordlevel, rdf:Value raccolta; for the item level rdf:Value brano to (4:BN)**
28	<ref>			dc:Reference (to 1:BN)	create 4:BN
29	#target	identification of the target resource		dc:identifier (to 4:BN)	
30	<unitdate>	date of the creation		dcterms:Created	crete instance of TimeSpan
31	#normal	normalised version		ens:CreatedNormal	create subclass of dc:Created called CreatedNormal
32	<geogname>	name of geografical location		dcterms:spatial (to 2:BN)	
33	<controlaccess>			/	
34	<geogname>	geographical location		dcterms:spatial	create instance of ens:Place, retrieve appropriate URIs for each toponym

35	#role	3 values: place, region, state	localita, regione, stato	&	create instance of ens:Place for each of the attribute values, replace literals with GeoName's appropriate URIs
36	<persname>	name of the performer /s		dc:Contributor	create instance of ens:Agent,(if the value of this element holds more than one name, extract them and create instance of ens:Agent for each)
37	#role	performer	esecuzione	dc:type (to its ens:Agent)	
38	#rules	voci miste, short description e.g.:2 mail voices and a guitar		dc:description (to Proxy)	
39	<subject>			dc:subject	
40	<materialspec>	information specific for the material		dc:abstract	create 5:BN
41	#type	"durata" -duration info "incipit"-starting lyrics	incipit/ durata		use only "incipit" value, ignore "durata" as already encoded in <metadigit>
42	<name>	serenata, carnevale		dc:subject	
43	#role	tradizione, ocasione		&	create skos:Concept for each of att values, i.e skos:Concept "tradizione", skos:Concept "ocasione"
44					
45	<descgrp>	information about materials having relationship to the unit described		ens:IsRelatedto	create 6:BN
46	#encodinganalog		ISAD 5 Allied materials area	dc:type (to 6:BN)	
47	<originalsloc>	info about the original support materials		/	
48	#type	support material	supporti	dc:type (to 7:BN)	
49	<item>	physical media in which resource is realised		dc:description	create 7:BN
50	<relatedmaterial>			/	
51	#type	annexes that go with the resource	allegati	dc:type (to 8:BN)	
52	<item>	fogli di registrazione RAI		dc:description	create 8:BN
53	<scopecontent>	transcribes of the lyrics!		dc:Abstract	create 9:BN
54	#encodinganalog		originale/traduzione	*	
55	<note>			dc:description	
56	#encodinganalog		ISAD 6 - 1 note	/	
57	<metadigit>	technical information about the MP3 version of recording		ens:IsRepresentationOf (to Proxy)	create instance 1 of InformationResource
58	<audio>			/	
59	<nomenclature>	same information as <unittitle>		/	
60	<sequence_number>	same information as <num>		/	
61	<proxies>			/	
62	<audio_dimension>			/	
63	<duration>	1' 38"		dcterms:extent (to 1:InformationResource)	
64	<audio_metrics>			/	
65	<samplingfrequency>	41 kHZ		/	
66	<bitrate>			/	
67	<file>			/	
68	#href			dcterms:hasFormat	
69	<format>			/	
70	<name>	MP3		dcterms:conformsTo (to 1:InformationResource	
71	<mime>	audio/MP3		dc:type (to 1:InformationResource)	render SOUND
72	<channel_configurati			/	
73	<compression>			/	
74	<usage>	conservatorio…		dc:description (to 9:BN)	
75	<dao>	digital archival object		/	
76	#audience		internal/external	the upper nod InformationResource is to be made only IF the value here is "external"	
77	#href			/	
78	#title			/	
79	<odd>			/	
80	#type	fonte_titolo			
81	<note>			dc:Description	

82	<daogrp>								
83		<daoloc>						/	
84			#title			repeats the information of the title		/	
formating elements:				<p>	<list> ->	copy as XMLLiterals			

Mapping Table 2; AV FOND (Fond "Audio Video")

Legend:
- (round brackets) hold the domain of the property, default property is ore:Poxy of the node;
- "/" no mapping;
- "~" same mapping as for the same element above, in these cases only the main elements are displayed without their sub-elements and;
- "->" look right;
- "&" see example 2, p...;
- "*" see example 1, p. ...;
-"^" see example 3, p. ..
-"#" attribute

	a) ANSC EAD ELEMENTS # -attribute	b) EXAMPLES OF SEMANTIC OF ANSC EAD ELEMENTS	c) FIXED VALUE OF THE ATT	d) EDM MAPPING ELEMENTS	e) ADDITIONAL WORK ON MAPPING : creation of objects, addidional comments; _:BN-blank node
1	<c>	fonds, series, subfonds		create instances of ens:ArchivalFond, ens:ArchivalSeries and ens:ArchivalSubfonds	create subclass of ens:NonInformationResource called ens:ArchivalFond , ens:ArchivalSeries and ens:ArchivalSubfonds
2	#level	fonds, series, subfonds		dc:type	
3	#id	identifier		dc:identifier	
4	#audience	internal/external		/	
5	<did>			/	
6	<unittitle>			/	
7	<title>			dc:title	Create 1:BN
8	#type		titolo suporto	dc:type (1:BN)	
9	<descgrp>	clustering element, no semantic		/	
10	#encodinganalog		ISAD 4	/	
11	<accessrestrict>			dc:rights	
12	#encodinganalog		ISAD 4-2	dc:type (2:BN)	
13	<c>	file		ore:Aggregation	
14	#level	file		dc:type	
15	#id	identifier		dc:identifier	
16	#audience		internal/external	/	
17	<did>			/	
18	<unittitle>			/	
19	<title>			dc:title	Create 2:BN
20	#type	support title…		*	
21	#authfilenumber	"" usually empty		^ (to 2:BN)	create a sub class of dc:Indentifier, called ens:authfilenumb
22	#source	the source of authfilenumber	AuthorityTitoliASC	dc:source (to 2:BN)	
23	<geogname>	IT, AT		dcterms:spatial; ens:happenedAt (event)	create instance 1 of Place
24	#authorityfilemnumber			ens:authfilenumber (to 1:Place)	
25	#role	nazione ,luogo, localita	nazione ,luogo, localita	&	new properties have ens:Place as domain

26	<bibref>	Reference on published material		ens:realizes	create instance 1 of ens:InformationResource (this InformationResource **isRepresentationOf** ore:Proxy for the nod aggregation
27	#role	information about publication	"dati sulla pubblicazione"	dc:description (to 1:InformationResource)	
28	#encodinganalog	not familiar with what does this stand for	500	/	
29	<name>	title, opera…		dc:description (to 1:InformationResource)	
30	#role	varianti, forma,		*	if #role is used map as dc:Description
31	#rules	Alternative Title	Altro titolo diverso del proprio tit	dcterms:alternative (to 1:InformationResource)	if #rules is used, map as dcterms:alternative
32	#type	Academic Title	Titilo Academico	dc:title (to 1:InformationResource)	create 3:BN; if #type was used, map as dc:title
33	<title>			rdf:value (to 3:BN)	
34	<num>	no actual number, no nextInSequence at this level		dc:description (to 3:BN)	
35	#type	opera..		dc:type (to 3:BN)	
36	#authfilenumber	authfilenumb		ens:authfilenum (to 3:BN) (^)	
37	<langmat>			/	
38	#label	A display label for an element	lingua	/	
39	<language>			dc:language (to 1:InformationResource)	
40	<imprint>	Information relating to the publication or distribution of a work cited in <bibref>		/	
41	<geogname>	Salburgo Festival… IT		~	
42	<date>			dcterms:issued (to 1:InformationResource)	
43	<publisher>			dc:Publisher (to 1:InformationResource)	
44	<unitdate>			/	
45	#normal			&	new property ens:normal has ens:TimeSpan as domain
46	<date>			dcterms:created, ens:OccuredAt (to Event)	create instance of ens:TimeSpan
47	<bibseries>	name of the series it belongs to		dc:source	
48	<origination>			/	
49	<persname>			dc:contributor; ens:HasMet (to Event)	instance of Agent create (ore:Proxy-1) that is proxyIn Event Aggregation
50	#altrender	alternative rendering		/	
51	#authfilenumber			ens:authfilenumber (^)	
52	#label	A display label for an element		/	
53	#role	autore, compositore		dc:type (to ore:Proxy-1)	
54	<emph>	instrument the person played of the voice		dc:description (to ore:Proxy-1)	
55	<corpname>	name of e.g. philharmonic orchestra		dc:contributor	instance of Agent create (ore:Proxy-2) that is proxyIn Event Aggregation
56	#authfilenumber			dc:identifier ens:HasMet (to Event)	
57	#label	A display label for an element		/	
58	#role	esecutore, autore		dc:type (to ore:Proxy-2)	
59	<geogname>			dcterms:spatial; ens:happenedAt (to Event)	
60	<physdesc>			dc:description	create 5:BN

61	#label		video, sonoro	dc:type (to 5:BN)	if existing, render SOUND of VIDEO, depending on value
62	#type	e.g. disco digitale		dc:type (to 5:BN)	
63	<physfacet>	elettrica/analogica		dc:type (to 5:BN)	
64	#type	configurazione		/	
65	<genreform>	originale, file audio		dc:type (to 5:BN)	
66	# type		WAVE	/	
67	<extent>			/	
68	# type	bitresolution, KHz		/	
69	<dymensions>			dcterms:extent (to 5:BN)	
70	# type		lungezza, largezza	/	
71	# unit		hh.mm.ss	/	
72	<archref>	no pcdata		/	
73	<note>	note on the material		dc:description (to 5:BN)	
74	<emph>	dall vivo , classica, non edito..		rdf:value (to 5:BN)	
75	<physloc>			/	
76	#id	physical location identifier		/	
77	<unitdate>			~	
78	<unitid>	empty, only attribute values		ens:currentLocation	instance 2 of ens:Place
79	#countrycode		IT	ens:country (to 2:Place)	
80	#label			/	
81	#repositorycode		ANSC	ens:provider (to 2:Place)	create instance of class:Agent, skosaltlable:ANSC, this URI will hold all the data on ANSC, address..+ skos altlable ANSC
82	#type	inventario, catalogo		/	
83	<repository>	Accademia Nazionale di Santa Cecilia		->	skos:prefLabel of URI describing provider Agent
84	#label		RM0	/	
85	<expan>			/	
86	#abbr		SC	->	skos:AltLabel for URI above
87	<abstract>			dc:abstract	
88	<materialspec>	information specific for this material		dc:description	
89	#label		registrazione, genre, TypeOf Reso	*	
90	<phystech>	physical condition e.g buono		/	
91	#encodinganalog	not familiar with encoding	CO-RS	/	
92	<note>			dc:description	
93	<controlaccess>			/	
94	<name>				
95	#role	event		ens:wasPresentAt (to Agent, InfResource, PhisThing of one nod)	create instance of EVENT , create an ore:Aggregation for Event
96	<date>			dcterms:creation; ens:occurredAt (to Event)	create instance of ens:TimeSpan

97	#certainity	/		/	
98	#normal			ens:normal (to ens:TimeSpan)	
99	#time	time of the beginning e.g. performance		ens:time (to ens:TimeSpan)	
100	<event>			/	
101	#id	identifier of event		dc:identifier (to Event)	
102	<num>	The name of the Series it belongs to e.g. "Stagione Sinfonica 1991-1992"		dc:source (to Event)	6:BN
103	#id	identifier of series		dc:identifier (6:BN)	
104	<subject>	interviste, pianoforte, 19 secolo…		dc:subject	
105	<persname>	person involved in event, already mapped		~	
106	<descgrp>	wrapper		/	
107	#encodinganalog	ISAD 3 content and structure area		/	
108	< scopecontent>	annexes of material described		dc:description	7:BN
109	#altrender	ISAD 3-1 scope and content	allegati	dc:type (to 7:BN)	
110	<accessrestrict>	no #PCDATA		/	
111	<legalstatus>	private property etc.		dc:rights	
112	<userestrict>			dc:rights	8:BN
113	#type	legalstatus, copyright..		dc:type (to 8:BN)	
114	<acquinfo>			dcterms:provenance	
115	<custodhist>			dcterms:provenance	
116	<metadigit>			->	ens:IsRepresentationOf (to Proxy), create instance 1 of ens:InformationResource
117	<video> <audio>	either <VIDEO> or <AUDIO> can occure		/	
118	<nomenclature>	title		/	
119	<proxies>			/	
120	<md5>	digital code, e.g. ba9fc080454321372a131d5eb23dd86d		/	
121	<usage>	1;b not familiar with value		/	
122	<video_dimension>			/	
123	<duration>	0:03:50		dcterms:extent (to 1:InformationResource)	
124	<video_metrics>			/	
125	<aspectratio>			/	
126	<framerate>			/	
127	<digitisation>			/	
128	<transcriptionagency>			/	
129	<devicesource>			/	
130	<sourcetype>			/	
131	<transcriptionchain>			/	
132	<capture_software>			/	
133	<device_description>			/	
134	#type			/	
135	<device_manufacturer>			/	

136							<device_model>				/	
137								#model			/	
138							<device_settings>				/	
139							<samplingfrequency>		44.1 KHz		/	
140						<transcription>					/	
141							<sourcetype>		Registrazione digitale audio		/	
142							<transcriptionagency>				/	
143							<transcriptionchain>				/	
144						<file>					/	
145							#href				dcterms:hasFormat	
146						<filesize>			53 mb		/	
147						<format>					dc:format (to 1:InformationResource)	
148							<mime>				dc:type (to 1:InformationResource)	
149							<name>		MP3		dcterms:conformsTo (to 1:InformationResource)	
150							<encode>				/	
151							<channel_configuration>				/	
152							<compression>				/	
153							<codec>		DivX 6.0		dcterms:isRequiredBy (to 1:InformationResource)	
154							<encode>		interlaced		/	
155							<videoformat>		PAL		dcterms:conformsTo (to 1:InformationResource)	
156						<audio_dimension>			0:03:50		/	
157							<duration>				dcterms:extent (to 1:InformationResource)	
158						<audio_metrics>					/	
159							<bitpersample>		24 bit		/	
160				<sequence_num>							/	indicator for making ens:NextinSequence with the previous element at the same level, if the value is sequential.
161	<c>								item level			
162		#level								item	create ore:Aggregation	
163		#id									~	
164		#audience									~	
165			<did>								/	
166				<unittitle>							dc:title	create 10:BN
167					<title>					titolo brano/titolo suporto	rdf:value for "titolo brano" (to 10:BN), dcterms:alternative for"titolo suporto"	
168						#type						
169						<name>			Composizione, Appassionata		rdf:value (to 10:BN)	
170						<num>			number of the track		dc:description (to 10:BN)	indicator for making ens:NextinSequence with the previous element at the same level, if the number value is sequential.
171					<bibref>						~	
172						<title>					~	
173						<name>					~	
174						<num>			Italienische		dc:description (to bibref mapping)	

175	<emph>				
176	<bibseries>			~	
177	<origination>			~	
178	<unitid>			~	
179	<abstract>			~	
180	<physdesc>			~	
181	<date>			dcterms:created(to Proxy), ens:occuredAt(to E	create instance of ens:Timespan
182	<geoname>	Auditorium Parco della Musica, Sala Sinopoli..		dcterms:spatial ens:happenedAt (to Event)	create instance of ens:Place,
183	<materialspec>			~	
184	<note>			~	
185	<controlaccess>			~	
186	<metadigit>			~	

formating elements: <imprint>, <emph>, <chronlist>, <cronitem>, <list>, <head>, <item> -> copy as XMLLiterals